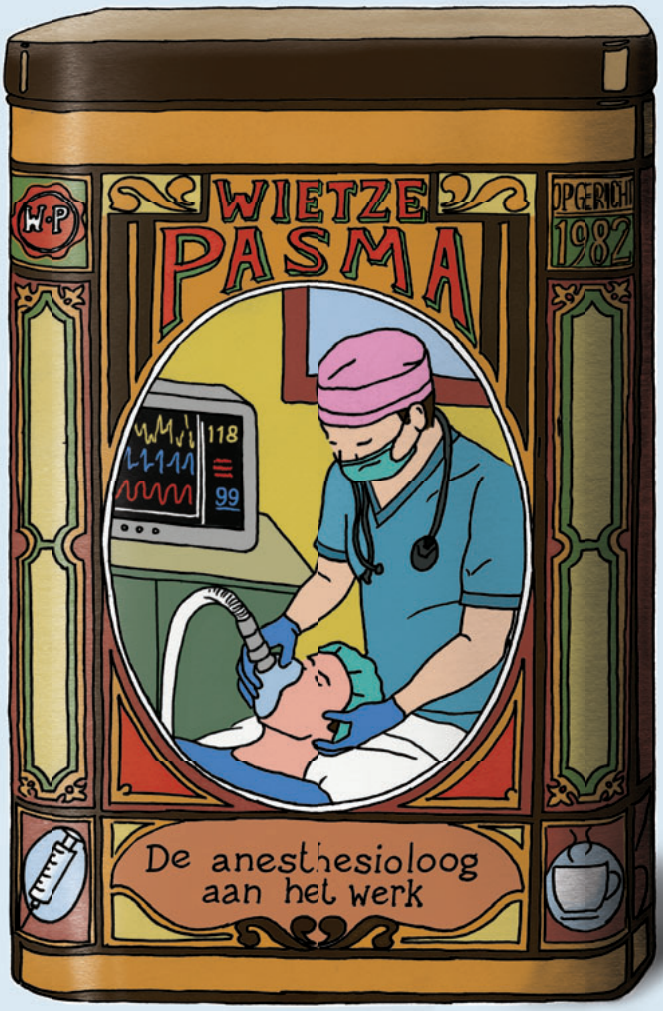
Use of anesthesia data for research



Use of anesthesia data for research

Wietze Pasma

W. Pasma
Use of anesthesia data for research
ISBN/EAN: 978-90-393-7298-2

Cover design by Annelieke Pasma, Pepijn Berghout (www.pepijnberghout.com) and Camiel Lemmens, persoonlijkproefschrift.nl
Layout and design by Camiel Lemmens, persoonlijkproefschrift.nl
Printing: Ridderprint | www.ridderprint.nl

Use of anesthesia data for research

Gebruik van anesthesiedata voor onderzoek

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 7 juli 2020 des middags te 12.45 uur

door

Wietze Pasma

geboren op 20 april 1982
te Hengelo (ov)

Promotoren:

Prof. dr. W.A. van Klei

Prof. dr. S. van Buuren

Copromotoren:

Dr. J.C. de Graaff

Dr. L.M. Peelen

Contents

1

General introduction

Anesthesia Information Management System (AIMS)

First there was pen and paper. Anesthesia health records were hand written, by surgeon or anesthesiologist in the operating room. The first records date back to the late 19th century.[1] The purpose of these charts was to get a better overview of peri-operative course of vital signs of the patient, and were not designed with research in mind[2]

To improve anesthesia documentation, anesthesiologists started experimenting with computers which could automatically capture physiological data from patient monitor hardware.[3] Built upon this principle, Anesthesia Information Management Systems (AIMS) were developed and implemented, starting from the late 1980s.[4] Besides physiologic parameters other events and medication administration were electronically documented in the AIMS, to get a complete overview of the anesthesia procedure. These electronic records were automatically capturing data resulting in more complete and better documentation than their analog counterparts.[5–9] Because data were already electronically available reusing these data for research purposes became much easier.[10]

AIMS were usually stand-alone systems, which did not communicate with other electronic health record systems. Nowadays these systems can be imbedded in or connected to the hospital electronic health record (EHR), providing a complete digital health record to an anesthesiologist and researcher. AIMS have not only improved anesthesia record keeping but also improved the academic output of anesthesia departments all over the world, such as in the field of blood pressure management and adverse outcomes such as myocardial injury and stroke.[11,12]

AIMS and research

AIMS databases are suitable for research because they carry rich data, encompassing a clearly defined timespan (anesthesia during a surgical procedure). Within this time period, the data collected has a high density (measurements per minute) and other event data such as medication administration are highly accurate. Therefore the link between events during anesthesia and adverse outcomes after anesthesia can be studied in detail.[11–13] However, anesthesia is a very safe procedure and negative outcomes due to anesthesia can be rare.[14] Research regarding such rare outcomes in single center databases therefore has limited value.

If data from multiple centers are combined, these investigations become possible or can be done more efficiently. There are several examples of large anesthesia data consortiums, which use administrative data.[15] The main purpose of these databases has often been quality control rather than research. In 2011 the multicenter perioperative outcomes group (MPOG, Ann Arbor, MI, USA) was formed for research purposes specifically.[16] In MPOG the main strategy used to come to a collaborative platform, was to share all data from the AIMS from all participating centers, rather than collecting and sharing data for a single research question. All participators can use the MPOG database for research. Proposals for the MPOG database have to be accepted by the consortium, before they can be realized, ensuring a certain quality of the research performed.

The strength of this system is its flexibility. Once the database was generated, it could answer far more than just one research question. On the other hand, sharing all AIMS data requires a bigger investment of time and resources. The local AIMS data has to be transformed to the required strict data format and data concepts (metadata) needs to be translated to the MPOG concept dictionary. Furthermore, data quality is as good or bad as the quality of data in the AIMS of each contributor, where data can be missing or invalid. Alternatively, setting up a dedicated (prospective) data collection as an alternative could yield better data, since variables can be checked directly and missing data can be manually added. [17]

Artifacts

Since the introduction of AIMS the quality and availability of data has improved. Research has shown that electronic data is more accurate and reliable, than manually collected data.[18] Even though the quality is higher, automatically captured data is unfiltered data and will also contain information which does not represent the true state of the patient at the time of measurement. These erroneous physiologic data are called artifacts.[19,20]

Whereas the introduction of AIMS improved the efficiency of data collection for clinical research, a different problem has risen, i.e. how to deal with these artifacts? Because artifacts could potentially influence the results of a study, the way artifacts are handled should be reported when presenting the results.[20] The gold standard of artifact handling is manual artifact identification. But this

method is labor intensive, time consuming and potentially subjective.[21] Several other methods, which do not require manual work, have been developed to filter AIMS data for artifacts, such as threshold filters and moving median filter.[22] However not much is known about the accuracy of these methods or which effect they might have on the results of a study.

Outline of this dissertation

The studies described in this dissertation aim to illustrate the possibilities of anesthesia database research and to study the impact of artifacts in anesthesia data on scientific research.

Part I of this dissertation focuses on intraoperative blood pressure in children.

In **Chapter 2** we describe the development of reference values for intraoperative blood pressure in children. Although the measurement of blood pressure is standard care, a reference value of what is normal, was not available for children. With data from multiple centers (using the MPOG database), we were able to construct such reference values. In **Chapter 3** we used these reference curves to normalize blood pressure data from a single center. In this exploratory study we evaluated whether children with a relatively low blood pressure are different from those with normal blood pressure.

Part II of this dissertation focuses on erroneous data, i.e. artifacts in AIMS databases. We aim to identify the size of this problem, evaluate whether it results to issues in database research and try to formulate a solution to the problem.

In **Chapter 4** we studied the incidence of artifacts in physiological data in children undergoing anesthesia. This study was setup similarly as was previously done in adults.[16] Apart from studying the incidences we also studied the association of factors with the incidence of artifacts, such as phase of surgery, surgical specialty and age of the patient.

We hypothesized that artifacts could influence study results. In **Chapter 5**, we analyzed the relation between intraoperative hypotension and post-operative myocardial injury as a case study. We applied different types of artifact filtering methods, to see if and to which extent effect estimates of this association would

be affected. We hypothesized that filtering artifacts in different ways, would lead to differences in effect estimates.

As manual artifact identification can be cumbersome, typically automatic artifact filters are being developed and their performance is then analyzed by comparing the results to manual identification. However, manual identification could also be influenced by factors. For example it could matter who annotated the data (an anesthesiologist or a researcher) or at which moment the data was annotated (live during surgery or retrospectively). Therefore, in **Chapter 6** we compared different annotation strategies for artifactual data. We then modelled these different annotations with machine learning algorithms, to see if the process of artifact identification could be automated. We hypothesized that despite the different situations in which the training data was collected, the identification of artifacts could be automated in a generalizable way with a machine learning algorithm.

Finally, **Chapter 7** evaluates the implications of the findings and provides a general discussion on the use of anesthesia data for research and data based algorithm development and **Chapter 8** summarizes the findings of this dissertation.

References

1. Sundararaman L V., Desai SP: The Anesthesia Records of Harvey Cushing and Ernest Codman. Anesth Analg 2018; 126:322–9
2. Gravenstein JS: The uses of the anesthesia record. J Clin Monit 1989; 5:256–65
3. Block FE, Burton LW, Rafal MD, Burton K, Newey C, Dowell L, Klein FF, Davis DA, Harmel MH: The computer-based anesthetic monitors: the Duke Automatic Monitoring Equipment (DAME) system and the microDAME. J Clin Monit 1985; 1:30–51
4. Stonemetz J: Anesthesia information management systems marketplace and current vendors. Anesthesiol Clin 2011; 29:367–75
5. Schalkwyk JM Van, Lowes D, Frampton C, Merry AF: Does manual anaesthetic record capture remove clinically important data? Br J Anaesth 2011; 107:546–52
6. Lerou JGC, Dirksen R, Daele M van, Nijhuis GMM, Crul JF: Automated charting of physiological variables in anesthesia: A quantitative comparison of automated versus handwritten anesthesia records. J Clin Monit 1988; 4:37–47
7. Thrush DN: Are automated anesthesia records better? J Clin Anesth 1992; 4:386–9
8. Benson M, Junger A, Fuchs C, Quinzio L, Bottger S, Jost A, Uphus D, Hempelmann G: Using an Anesthesia Information Management System to prove a deficit in voluntary reporting of adverse events in a quality assurance program. J Clin Monit Comput 2000; 16:211–7
9. Edwards KE, Hagen SM, Hannam J, Kruger C, Yu R, Merry AF: A randomized comparison between records made with an anesthesia information management system and by hand, and evaluation of the Hawthorne effect. Can J Anesth 2013; 60:990–7
10. Goehner NS, Roman PEF: Monitoring Technologies in Acute Care Environments: A Comprehensive Guide to Patient Monitoring Technology. 2014
11. Waes JAR van, Klei WA van, Wijeysundera DN, Wolfswinkel L van, Lindsay TF, Beattie WS: Association between Intraoperative Hypotension and Myocardial Injury after Vascular Surgery. Anesthesiology 2016; 124:35–44
12. Study ANC, Bijker JB, Persoon S, Peelen LM, Moons KGM, Kalkman CJ, Kappelle LJ, Klei WA van, Study ANC: Intraoperative hypotension and perioperative ischemic stroke after general surgery: a nested case-control study. Anesthesiology 2012; 116:658–64

13. Bijker JB, Klei WA van, Vergouwe Y, Eleveld DJ, Wolfswinkel L van, Moons KGM, Kalkman CJ: Intraoperative hypotension and 1-year mortality after noncardiac surgery. Anesthesiology 2009; 111:1217–26
14. Amalberti R, Auroy Y, Berwick D, Barach P: Five system barriers to achieving ultrasafe health care. Ann Intern Med 2005; 142:756–64
15. Levin MA, Wanderer JP, Ehrenfeld JM: Data, big data, and metadata in anesthesiology. Anesth Analg 2015; 121:1661–7
16. Kheterpal S: Clinical research using an information system: The multicenter perioperative outcomes group. Anesthesiol Clin 2011; 29:377–88
17. Simpao AF, Pruitt EY, Cook-Sather SD, Gurnaney HG, Rehman MA: The reliability of manual reporting of clinical events in an anesthesia information management system (AIMS). J Clin Monit Comput 2012; 26:437–9
18. Edsall DW, Deshane P, Giles C, Dick D, Sloan B, Farrow J: Computerized patient anesthesia records: less time and better quality than manually produced anesthesia records. J Clin Anesth 1993; 5:275–83
19. Kool NP, Waes J a R van, Bijker JB, Peelen LM, Wolfswinkel L van, Graaff JC de, Klei W a van: Artifacts in research data obtained from an anesthesia information and management system. Can J Anaesth 2012; 59:833–41
20. Martin M, Kheterpal S: Monitoring artifacts and large database research: what you don't know could hurt you. Can J Anesth 2012; 59:823–6
21. Verduijn M, Peek N, Keizer NF de, Lieshout EJ van, Pont ACJM de, Schultz MJ, Jonge E de, Mol BAJM de: Individual and Joint Expert Judgments as Reference Standards in Artifact Detection. J Am Med Informatics Assoc 2008; 15:227–34
22. Cunningham S, Symon AG, McIntosh N: The practical management of artifact in computerised physiological data. Int J Clin Monit Comput 1994; 11:211–6

I

Blood pressure in pediatric anesthesia

2

Reference Values for Noninvasive Blood Pressure in Children during Anesthesia: A Multicentered Retrospective Observational Cohort Study

Wietze Pasma *
Jurgen C. de Graaff *
Stef van Buuren
Jesse J. Duijghuisen
Olubukola O. Nafiu
Sachin Kheterpal
Wilton A. van Klei

* contributed equally as co-first authors.

Published in Anesthesiology 2016 Nov;125(5):904-913

Abstract

Background

Although noninvasive blood pressure (NIBP) monitoring during anesthesia is a standard of care, reference ranges for blood pressure in anesthetized children are not available. We developed sex- and age-specific reference ranges for NIBP in children during anesthesia and surgery.

Methods

In this retrospective observational cohort study, we included NIBP data of children with no or mild comorbidity younger than 18 yr old from the Multicenter Perioperative Outcomes Group data set. Sex-specific percentiles of the NIBP values for age were developed and extrapolated into diagrams and reference tables representing the 50th percentile (0 SD), +1 SD, −1 SD, and the upper (+2 SD) and lower reference ranges (−2 SD).

Results

In total, 116,362 cases from 10 centers were available for the construction of NIBP age- and sex-specific reference curves. The 0 SD of the mean NIBP during anesthesia varied from 33 mmHg at birth to 67 mmHg at 18 yr. The low cutoff NIBP (2 SD below the 50th percentile) varied from 17 mmHg at birth to 47 mmHg at 18 yr old.

Conclusions

This is the first study to present reference ranges for blood pressure in children during anesthesia. These reference ranges based on the variation of values obtained in daily care in children during anesthesia could be used for rapid screening of changes in blood pressure during anesthesia and may provide a consistent reference for future blood pressure–related pediatric anesthesia research.

Introduction

More than six million patients under the age of 18, including 1.5 million infants, undergo procedures requiring sedation and anesthesia each year in the United States and Europe alone.[1] While concerns about potential neurotoxicity of general anesthesia have dominated the lay press and peer-reviewed literature, all pediatric patients undergoing surgery are at risk of complications related to anesthesia-induced cardiovascular depression. In fact, some have voiced concerns that neurotoxicity may be mediated by hypoperfusion.[2]

Monitoring of vital signs, such as heart rate and blood pressure, during anesthesia is a standard of care according to the American Society of Anesthesiologists (ASA) practice guidelines.[3] Although noninvasive blood pressure (NIBP) is obtained at least every 5 min in every patient, the interpretation of the values obtained in children is not guided by evidence because there are no valid reference ranges for NIBP in children during anesthesia. Current NIBP reference values are derived from population-based studies in healthy nonanesthetized children.[4–6] However, these references cannot be extrapolated to children during anesthesia because anesthetic drugs induce significant cardiovascular depression.[7]

This gap in knowledge causes uncertainty, in particular, due to the known developmental physiologic spectrum of the pediatric patient. A recent case series of six healthy infants who developed severe encephalopathy after straightforward elective surgery, including one death and one with profound developmental delay, has refocused attention on pediatric anesthesia standards of care.[8] The exact cause of the encephalopathy could not be determined in any of these children, but hypotension during anesthesia was considered attributable. However, provider experience demonstrates that many infants with comparably hypotensive NIBP during anesthesia are doing well after surgery and do not develop any complications. The absence of evidence-based reference ranges exposes patients to nonstandardized care and providers to undue medicolegal risk while providing routine, reasonable care.

The purpose of the current study was to develop age- and sex-specific charts (percentile curves) of the natural variation of NIBP during anesthesia during the preparation and surgical phase. Availability of such reference curves and tables of NIBP measurements would aid physicians in detecting extreme values that

justify further inquiry or action and would enable harmonization of diagnosis and treatment between anesthesiologists, hospitals, and countries.[9]

Materials and Methods

In the current retrospective observational cohort study, we used deidentified data from the Multicenter Perioperative Outcomes Group (MPOG) data set. The MPOG is a consortium of medical centers in the United States and Europe aggregating large volumes of observational perioperative electronic data, registry outcomes, and long-term administrative outcomes.[10] The study of deidentified patient data from the MPOG data set has been approved by the University of Michigan Institutional Review Board, Ann Arbor, Michigan (HUM00033894). Participation of the University Medical Center Utrecht (UMCU), Utrecht, The Netherlands, in the MPOG database has been approved by the Institutional Review Board of the UMCU (12.253-C). The protocol for the current study was approved by the MPOG Perioperative Clinical Research Committee (0018, February 10, 2014).

Data Source

The MPOG database has been described previously but is summarized as follows: each contributing member uses a modern intraoperative electronic health record (EHR). Data manually entered into the EHR include patient anthropometrics, intraoperative medications, surgical events, and clinical observations. Data collected using automated, validated interfaces include demographic information, laboratory values, and all physiologic monitor measurements, including heart rate and NIBP.[10] The database is growing over time and, at the time of data extraction, contained the automatically acquired data on vital signs and other EHR data of more than 3.5 million patients, including around 350,000 children from 17 centers. We included children with American Society of Anesthesia physical status (ASA PS) classes 1 and 2 who underwent anesthesia for surgical or diagnostic procedures in hospitals participating in the MPOG in the period from August 2007 up to and including December 2014. ASA PS 1 children are considered healthy, and ASA PS class 2 children have minor comorbidity not affecting daily life, such as controlled bronchial asthma.[11] We excluded cardiac surgery cases and cases for which the type of surgery was unknown.

Although NIBP measurements are routinely acquired at least every 5 min during anesthesia, only measurements acquired during relatively stable hemodynamic

periods were used to provide insight into reference values for NIBP. The analysis was therefore performed on data obtained during the preparation phase (within 20 min before procedure start) and during the initial surgical phase after the start of the procedure (between 15 and 35 min after procedure start). This was done in order to minimize the effect of patient positioning, initial surgical stimulus surge, and surgical bleeding on the reference values. Additionally, to estimate how often blood pressures during the remainder part of the surgical procedure were outside the range of the derived reference values, blood pressures obtained after the first 35 min were compared to the reference values.

We included only cases in which the start of the surgical procedure was clearly identified. For the preparation phase (20 min before procedure start), we calculated the mean of the last three measurements without artifacts, within this time frame. During the surgical phase, we calculated the mean of the first three measurements without artifacts, between 15 and 35 min after the start of the procedure. A case was included when at least two artifact-free measurements were available in the given time frame.

Although the management of artifact data is essential to provide meaningful reference ranges, no standardized definitions of blood pressure artifacts have been established for pediatric blood pressure measurements. Measurements were considered artifacts when the diastolic NIBP was lower than 3 mmHg, when the systolic NIBP was higher than or equal to 250 mmHg, when the pulse pressure (systolic NIBP minus diastolic NIBP) was lower than or equal to 5 mmHg, when the diastolic NIBP was higher than the mean NIBP, when the mean NIBP was higher than the systolic NIBP, or when the systolic, diastolic, or mean NIBP measurement was flagged as an artifact or was missing.

Outcomes

The primary outcome of the study was an estimate of the natural variation of systolic, mean, and diastolic NIBP in relation to age during the preparation and surgical phase of the procedure.[4] The variation was expressed as sex-specific percentile curves with the 50th percentile (0 SD), −1 SD, +1 SD, and highest (+2 SD) and lowest (−2 SD) reference ranges. Secondary outcomes were sex-specific plots of percentile curves and reference tables of systolic, mean, and diastolic NIBP in relation to weight.

Statistical Analysis

The NIBP values were analyzed according to the methods recommended by the World Health Organization for child growth standards.[12] For each outcome, a choice was made between the Box-Cox Cole-Green, the Box-Cox Power Exponential, and the Box-Cox t-family of distributions, aided by the worm plot and Q statistics as implemented in the generalized additive models for location, scale, and shape package for R software (version 3.2.3; https://www.r-project.org, accessed August 13, 2016).[13,14] Where needed, transformations of the variables were used to increase the fit. Differences between centers were estimated by adding dummy variables, as practiced in this context and evaluated with worm plots.[15] The results of systolic, diastolic, and mean NIBP are presented in graphs in relation to age and weight with a square root transformation of age for more detailed presentation for younger children. The R scripts to fit the models can be requested from the authors.

In order to compare the blood pressures during the remainder of the procedure with the proposed reference scales for age and weight, all NIBP measurements from 35 min after the start of the procedure until the end of anesthetic registration were gathered. These measurements were divided into epochs, each including three measurements. For each epoch, the median of the NIBP was calculated and compared with the preparation phase reference value of the same age or weight.

By including all surgical and diagnostic procedures on children under 18 yr old conducted between August 2007 and December 2014, we expected to include more than 100,000 cases. This sample is much larger than is customary in the field (typically n = 1,000 to 10,000).[12]

Results

Of a total of 327,123 anesthetics in patients younger than 18 yr old in the MPOG database, 221,202 cases with known ASA PS classes 1 and 2 and known sex were selected. Cardiac surgery cases, cases where the surgical procedure was unknown, and those without a documented “procedure start” timed event were excluded (figure 1). In total, 116,362 cases across 10 centers with at least two valid blood pressure measurements in any of the periods of interest were used for analyses (University of Michigan Health System, Ann Arbor, Michigan; Oregon Health and Science University, Portland, Oregon; University of Colorado Denver, Denver,

Colorado; University of Tennessee Medical Center, Knoxville, Tennessee; University of Virginia Health System, Charlottesville, Virginia; University of Florida, Gainesville, Florida; Washington University School of Medicine, St. Louis, Missouri; University of Vermont – Fletcher Allen Health Care, Burlington, Vermont; University of Washington Medical Center, Seattle, Washington; and University Medical Center Utrecht, Utrecht, The Netherlands). We used 108,179 cases for analysis of the preparation phase and 94,283 cases for analysis of the initial surgical phase. The age was known in all cases; in 104,977 cases (90%), the weight was known (table 1). The worm plots did not show significant differences between centers (data not presented).

Table 1: Baseline Characteristics of Included Pediatric Cases (n = 116,362)

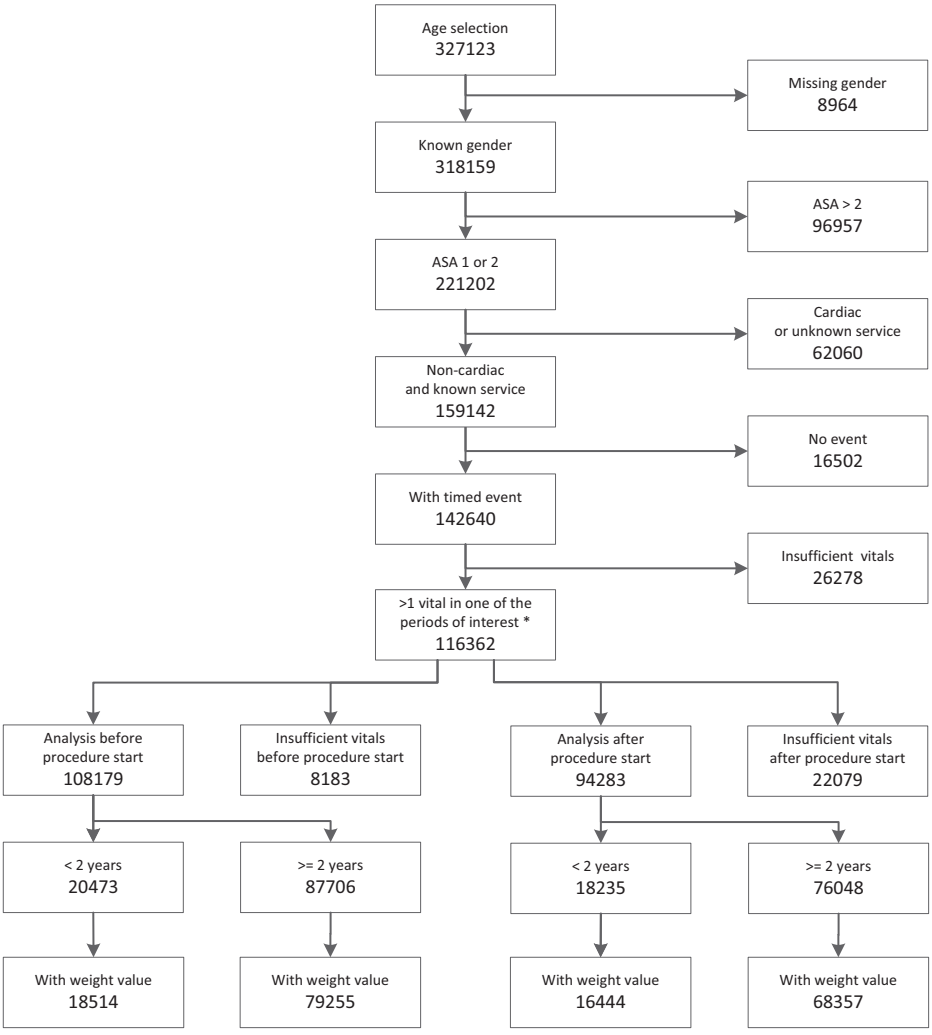
	All participants		**< 2 years old**	
Cases	116362		22455	
Age				
< 2 months	2122	(1.8%)	2122	(9.5%)
2 months – 12 months	11051	(9.5%)	11051	(49.2%)
1 year old	9282	(8.0%)	9282	(41.3%)
2 years – 5 years old	28359	(24.4%)		
6 years – 8 years old	16624	(14.3%)		
9 years – 11 years old	13584	(11.7%)		
12 years or older	35340	(30.4%)		
ASA				
1	59581	(51.2%)	12620	(56.2%)
2	56781	(48.8%)	9835	(43.8%)
Vital signs available				
Only pre-procedure start	22079	(19.0%)	4220	(18.8%)
Only post-procedure start	8183	(7.0%)	1982	(8.8%)
Both	86100	(74.0%)	16253	(72.4%)

Table 1 *(continued)*: Baseline Characteristics of Included Pediatric Cases (n = 116,362)

	All participants		< 2 years old	
Weight (median, IQR)	24.1	(14.0 - 50.0)	9.1	(6.9 - 10.9)
Weight known	104977	(90.2%)	20240	(90.1%)
Height (median, IQR)	127	(96 - 160)	73	(64 - 79)
Height known	78242	(67.2%)	13358	(59.5%)
Sex				
Male	67501	(58.0%)	15190	(67.6%)
Female	48861	(42.0%)	7265	(32.4%)
Surgical service				
Otolaryngology	21636	(18.6%)	3255	(14.5%)
Orthopedics	20153	(17.3%)	1246	(5.5%)
Urology	16473	(14.2%)	5871	(26.1%)
General	14577	(12.5%)	4513	(20.1%)
Ophthalmology	10489	(9.0%)	2140	(9.5%)
Plastics	7898	(6.8%)	2125	(9.5%)
Other	25136	(21.6%)	3305	(14.7%)

Values given in number of cases (n) and percentage of group (%), unless otherwise stated.

Figure 1: Flow diagram of data inclusion from Multicenter Perioperative Outcomes Group data set.



The 50th percentile (0 SD) of the systolic NIBP during anesthesia for boys varies from 48 mmHg at birth to 100 mmHg at 18 yr, while the 50th percentile of the mean NIBP during anesthesia varies from 33 to 66 mmHg, respectively (figure 2 and table 1 in Supplemental Digital Material). The reference value of the mean NIBP in the lowest age range in the preparation phase varies between 17 mmHg (−2 SD) and 71 mmHg (+2 SD) for boys and between 18 mmHg (−2 SD) and 59 mmHg (+2 SD) for girls (table 2 in Supplemental Material).

Sex-specific percentile plots of systolic, mean, and diastolic NIBP are presented in relation to age during the preparation (figure 2) and surgical phase (figure 3) with associated reference tables (preparation phase: tables 1 to 3 in Supplemental Material, and surgical phase: tables 4 to 6 in Supplemental Material). The sex-specific percentile plots of systolic, mean, and diastolic NIBP in relation to weight are presented in figures 4 and 5 and tables 7 to 12 in Supplemental Material. Detailed tables in relation to age, weight, and height are available on request from the corresponding author.

In 62,727 (58%) of the 108,179 cases, at least one epoch of three blood pressure measurements was available after the initial 35 min of the procedure start, i.e., in these 62,727 cases, the procedure lasted longer than 35 min. In total, 405,417 epochs were identified, and the median NIBP of these epochs was compared to the corresponding reference value of NIBP. This resulted in 1.25% of the NIBP below the corresponding −2 SD and 1.79% of the NIBP above the 2 SD. Weight was known in 56,643 (90%) of these 62,727 cases. Similar comparison of NIBPs resulted in 1.34% and 1.24% epoch NIBPs below the −2 SD and 1.68% epoch NIBPs above the 2 SD using the weight-adjusted reference curves.

Figure 2: Reference curves for noninvasive blood pressure for boys and girls during anesthesia during the preparation phase in relation to age.

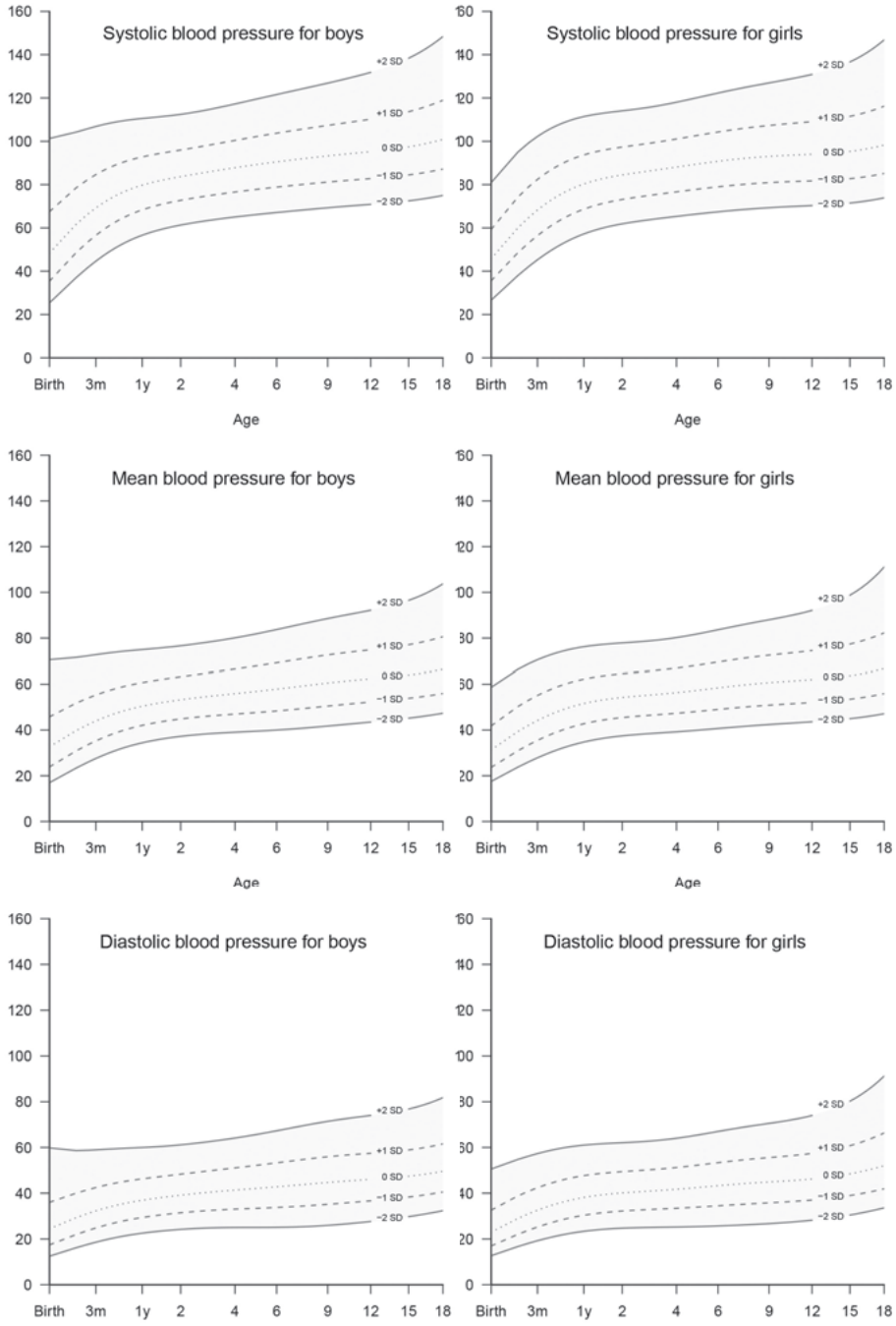


Figure 3: Reference curves for noninvasive blood pressure for boys and girls during anesthesia during the surgical phase in relation to age.

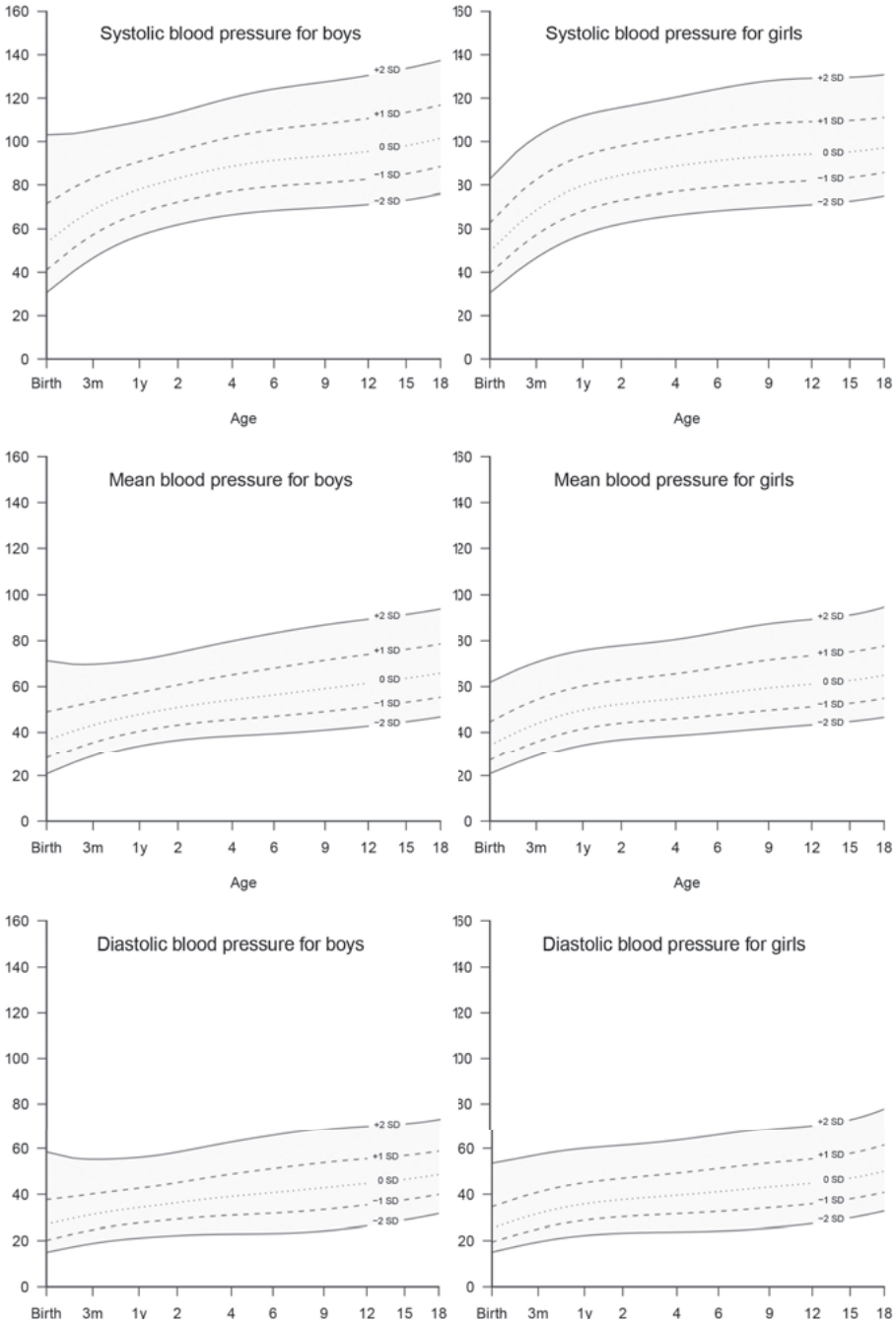


Figure 4: Reference curves for noninvasive blood pressure for boys and girls during anesthesia during the preparation phase in relation to weight.

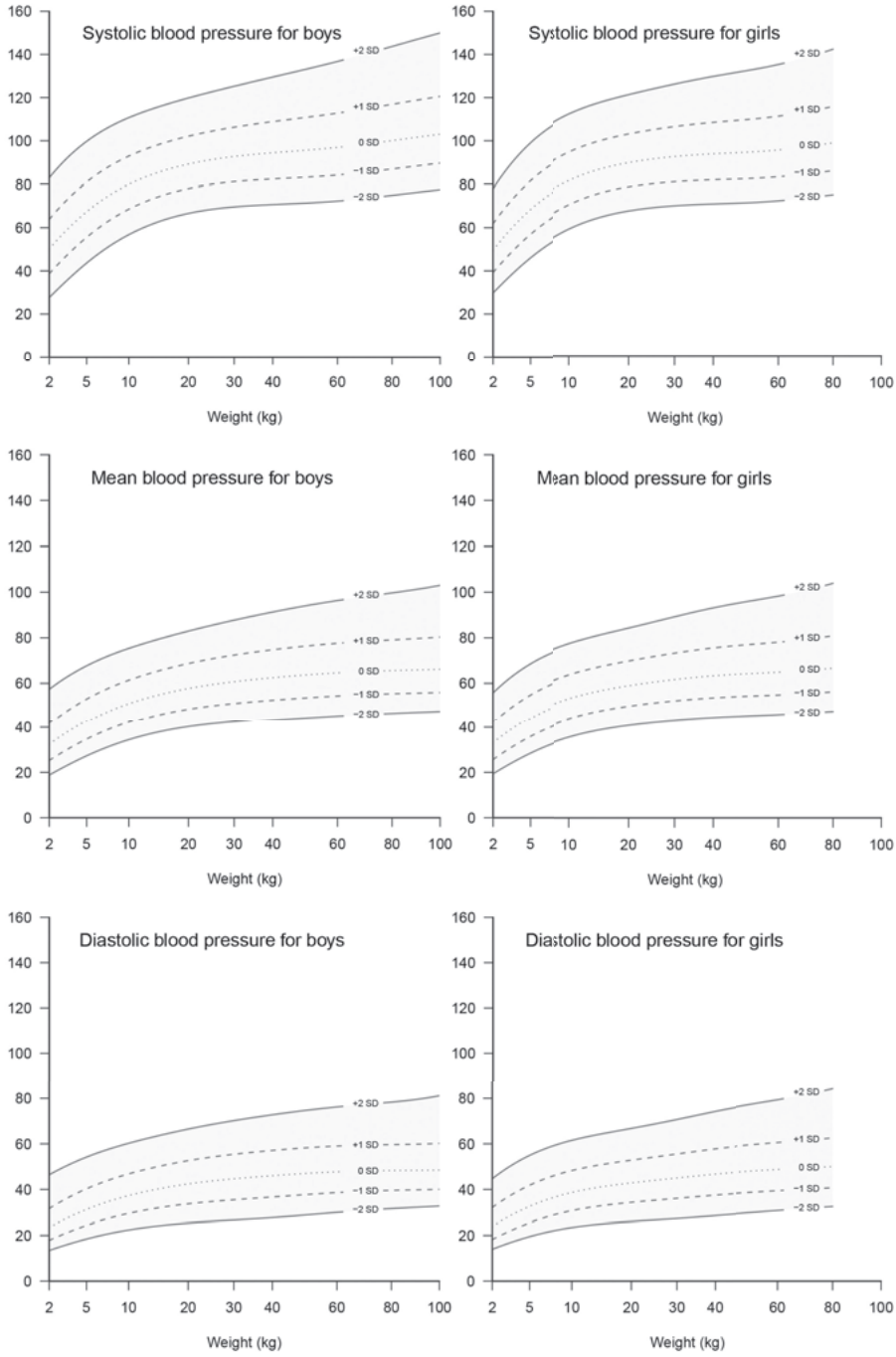
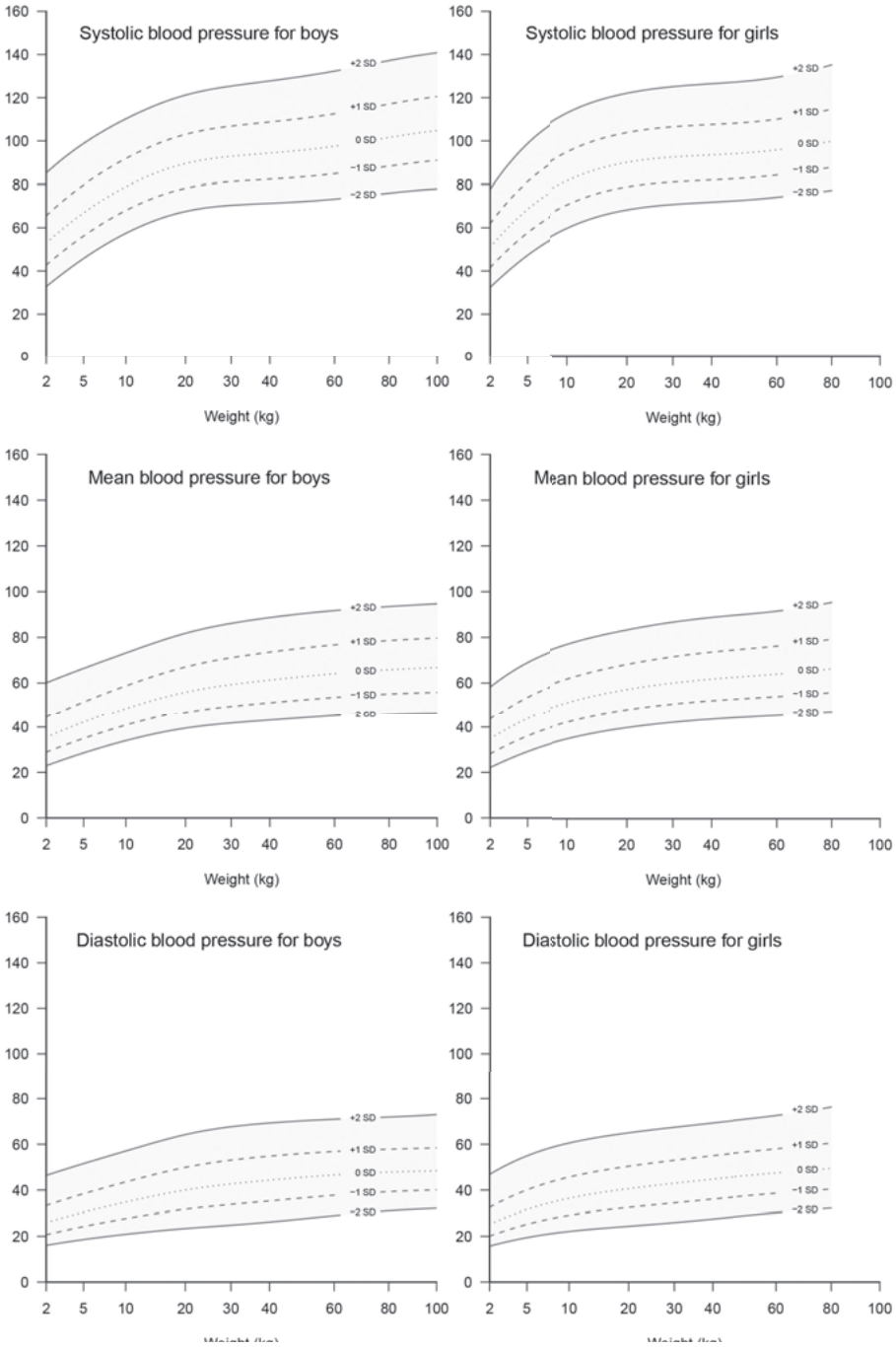


Figure 5: Reference curves for noninvasive blood pressure for boys and girls during anesthesia during the surgical phase in relation to weight.



Discussion

Frequent monitoring of NIBP during anesthesia is a universal standard of care. However, reference values for NIBP in children during anesthesia did not previously exist. The present multicenter study produced age- and sex-specific NIBP reference values for children during anesthesia. The 50th percentile (0 SD) of the mean NIBP during the surgical preparation phase varies from 33 mmHg at birth to 66 mmHg at 18 yr during anesthesia for boys. The lower reference values in anesthetized children are considerably lower than those in nonanesthetized children. The −2 SD of the mean NIBP varies between 17 mmHg for the youngest and 47 mmHg for children of 18 yr old. More than 95% of the blood pressure values of the remaining part of the surgical procedure was within the obtained reference ranges, i.e., values were above the −2 SD and below the +2 SD.

Lack of Knowledge

Thus far, reference ranges for NIBP in children during anesthesia do not exist. Vital signs during anesthesia, in particular, NIBP measurements in adults, are often evaluated in reference to patient-specific baseline measurements. However, the clinical relevance of specific NIBP targets or thresholds is disputable, and even for adults, a clear definition of intraoperative hypotension does not exist.[16] The absence of reference values for NIBP during anesthesia causes a wide variation in diagnosis of potential derangement and subsequent treatment. For example, the threshold for intraoperative hypotension in children that is used in daily practice varies between anesthesiologists: some use a threshold at a 10% decrease from the baseline value, while others use a 40% decrease.[17] Furthermore, acceptable patient-specific baseline measurements are rarely available in children since obtaining a reliable NIBP measurement in uncooperative children in the operating room before anesthesia induction is almost impossible. This results in large discrepancies between definitions of intraoperative hypotension. Pediatric anesthesiologists from the Society of Pediatric Anesthesia and Association of Paediatric Anaesthetists reported definitions of intraoperative hypotension ranging from a systolic blood pressure of 25 to 70 mmHg for neonates and from 40 to 100 mmHg in children between 2 and 12 yr old.[17] This wide variation in clinician decision-making demands generalizable, multicenter data to inform practice.

Interpretation

The interpretation of the reference ranges of NIBP presented in our study is comparable to the reference ranges of growth standards for children presented by the World Health Organization, in which measurements of weight of a whole population are plotted into percentile curves.[12] Customarily, observations outside the 2.5th (−2 SD) and 97.5th (+2 SD) percentiles are considered unusually low or high.[12,18] The goal of the reference ranges presented is to provide a set of reference values for the practitioner enabling him or her to judge measurements acquired during clinical practice in relation to the variation within the reference population. The present reference curves should not be related to outcome, as they represent the variation in the population, as is the current practice with growth charts. An unusually low or high value relative to the reference alerts the medical practitioner, who can judge the measurement in the context of the clinical setting, patients' disease, the specific anesthetic medication, and the phase and type of the procedure for which the anesthetic is being administered after which appropriate treatment can be started, if judged necessary.[4]

In nonanesthetized children, reference values are focused on hypertension and are defined in relation to age and sex.[19] Therefore, most reference tables for nonanesthetized children focus on the upper limits (+2 SD) of the reference values.[6] The upper reference range of the systolic NIBP differs only marginally between nonanesthetized and anesthetized children. The 98th percentiles of the systolic NIBP of boys 4 and 10 yr old when not anesthetized are 120 and 126 mmHg, respectively,[6] and are comparable to our observed results: systolic NIBP when anesthetized during the preparation phase (+2 SD: 117 and 128 mmHg, respectively) and during the surgical phase (+2 SD: 120 and 130 mmHg, respectively). In contrast, the 50th percentile (0 SD) and 2.5th percentile (−2 SD) NIBPs are much lower in anesthetized compared to nonanesthetized children. Published guidelines establish that the 50th percentiles (0 SD) of the systolic NIBP of boys 4 and 10 yr old when nonanesthetized are 103 and 108 mmHg, respectively.[6] However, our real-world data demonstrate that 50th percentile (0 SD) for the same type of patients during the preparation phase of surgery is approximately 15 mmHg lower (88 and 94 mmHg, respectively). These figures are 90 and 95 mmHg, respectively, during the surgical phase. The lowest reference ranges for anesthetized children are considerably lower (approximately 20 mmHg) than those for awake children; the 2.5th percentiles (−2 SD) of the systolic and mean NIBP of a 4-yr-old boy are 85

and 60 mmHg, respectively, when nonanesthetized,[17] and 68 and 38 mmHg, respectively, when anesthetized.

The variance of NIBPs found is rather large. The wide range is caused by the natural biologic variation, differential impact of anesthesia medication on cardiovascular depression, and variation in the measurement technique itself (i.e., location, the circumference of limb being measured, the cuff size used, etc.). These confounding factors are not reliably available in the present multicenter retrospective cohort study data and therefore cannot be used to adjust the model. In contrast, the NIBP reference values in awake children are based on highly controlled ambulatory care settings enabling tight measurement criteria: auscultatory measurements, on the right arm, using appropriate cuff size (an inflatable bladder width that is at least 40% of the arm circumference at a point midway between the olecranon and the acromion) in a sitting position.[19] These ideal and systematic measurements are seldom possible in the operating room. In the current anesthesia standard of practice, an automatic oscillometric device is used, which measures mean arterial blood pressure and then calculates the systolic and diastolic values,[20] whereas the variation may be increased by the algorithms used for the devices, which may differ from company to company.[21] Furthermore, the location of the NIBP varies among both arms and legs, due to the placement of intravenous lines and surgical site. When every blood pressure value in this study was measured under ideal circumstances, the variability might have been smaller, but this ideal measurement circumstances generally are not present during anesthesia. Nevertheless, this wide variance represents the natural variation each pediatrician and anesthesiologist has to deal with in daily practice.

The current reference ranges have a wide variation in the lower end of the spectrum of age and weight ranges. This wide variation reflects the variation in clinical practice and is not caused by excluding incomplete data since the reasons for including or excluding patients were not dependent on blood pressure. Therefore, analysis of the complete cases will not introduce bias in the estimated distributions.[22] We excluded severely diseased children (ASA PS 3 and higher) to limit the influence of these sicker children on the blood pressure reference to focus on the initial research question (healthy children).

To give insight into the pressure ranges during the remaining part of the procedures, we compared these blood pressures with the reference standards proposed here. These results show that more than 95% of these measurements were within these limits, and only a very limited number of measurements were outside the reference range. Short procedures are not represented in this last analysis because of the method of data collection, which could have led to under- or overestimation of the percentages.

Organ Perfusion

The present data set does not allow a comparison of NIBP in relation to the clinical outcome related to organ perfusion, such as cerebral encephalopathy, kidney failure, or even surrogate outcome measures, such as hospital length of stay and mortality. Clear symptoms are very rare in these group of patients, and if present, they would not affect the results of the current study.

A cerebral ischemic infarct is a rare but devastating complication in pediatric anesthesia caused by low cerebral perfusion.[8] The symptoms of low cerebral perfusion (syncope, nausea, etc.) are not signaled during general anesthesia. The lowest systolic NIBP of the patients with severe postoperative encephalopathy in the case series by McCann et al.[8] ranged from 22 to 40 mmHg. The current study shows considerably lower reference ranges (–2 SD) of the systolic and mean NIBP in the youngest boys of 25 and 17 mmHg during the preparation phase and 30 and 22 mmHg during the surgical phase. This lower range (–2 SD) corresponds with the lower 10th percentiles of the definition of systolic NIBP for significant hypotension (25 mmHg) that pediatric anesthesiologists from the United States and United Kingdom reported.[17] These differences underscore the variation and uncertainty of current clinical practice.

Conclusions

Clinicians providing care to millions of pediatric patients during anesthesia each year lack the normative data needed to guide interpretation and treatment of intraoperative NIBP measurements. With more than 100,000 patients across 11 medical centers, our data provide the real-world context necessary to establish reference ranges for a pediatric anesthesia blood pressure. We have demonstrated that the 50th percentile (0SD) and lower bound (–2 SD) NIBPs observed during anesthesia are markedly lower in anesthetized children than in idealized data

collected on awake patients. These data may help inform daily practice and decision-making for thousands of clinicians—anesthesiologists, surgeons, pediatricians, and pediatric specialists. Furthermore, the external validity of studies of anesthetized children can be calibrated to these reference tables.

Acknowledgments

The authors acknowledge the faculty representatives of the Multicenter Perioperative Outcomes Group Perioperative Clinical Research Committee from the following sites: Oregon Health and Science University, Portland, Oregon; Columbia University, New York, New York; University of Oklahoma, Oklahoma City, Oklahoma; University of Colorado, Denver, Colorado; University of Utah, Salt Lake City, Utah; University of Tennessee Medical Center, Knoxville, Tennessee; University of Virginia Health System, Charlottesville, Virginia; University of Florida, Gainesville, Florida; Washington University School of Medicine, St. Louis, Missouri; University of Vermont – Fletcher Allen Health Care, Burlington, Vermont; University of Washington Medical Center, Seattle, Washington; University of Michigan Health System, Ann Arbor, Michigan; and University Medical Center Utrecht, The Netherlands.

Research Support

Supported by the Christine Bader Foundation Irene Children's Hospital (Christine Bader Stichting Irene Kinderziekenhuis, Arnhem, The Netherlands), an unrestricted research grant from the European Society of Anesthesiology (Brussels, Belgium) 2015 sponsored by Philips (Amsterdam, The Netherlands), and departmental resources. The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Competing Interests

The work of Dr. Nafiu is in part supported by grant K23 GM104354 from the National Institute of General Medical Sciences (Bethesda, Maryland). The other authors declare no competing interests.

References

1. DeFrances, CJ, Cullen, KA, Kozak, LJ National Hospital Discharge Survey: 2005 annual summary with detailed diagnosis and procedure data.. Vital Health Stat 13 . (2007). 1–209
2. McCann, ME, Soriano, SG Perioperative central nervous system injury in neonates.. Br J Anaesth. (2012). 109suppl 1 i60–7
3. Available at: http://www.asahq.org/quality-and-practice-management/standards-and-guidelines. Accessed July 22, 2016
4. Blake, KV, Gurrin, LC, Evans, SF, Newnham, JP, Landau, LI, Stanley, FJ, Beilin, LJ Reference ranges for blood pressure in preschool Australians, obtained by oscillometry.. J Paediatr Child Health. (2000). 36 41–6
5. Neuhauser, HK, Thamm, M, Ellert, U, Hense, HW, Rosario, AS Blood pressure percentiles by age and height from nonoverweight children and adolescents in Germany.. Pediatrics. (2011). 127 e978–88
6. Ingelfinger, JR Clinical practice. The child or adolescent with elevated blood pressure.. N Engl J Med. (2014). 370 2316–25
7. Lerman, J, Davis, PJ, Welborn, LG, Orr, RJ, Rabb, M, Carpenter, R, Motoyama, E, Hannallah, R, Haberkern, CM Induction, recovery, and safety characteristics of sevoflurane in children undergoing ambulatory surgery. A comparison with halothane.. Anesthesiology. (1996). 84 1332–40
8. McCann, ME, Schouten, AN, Dobija, N, Munoz, C, Stephenson, L, Poussaint, TY, Kalkman, CJ, Hickey, PR, de Vries, LS, Tasker, RC Infantile postoperative encephalopathy: Perioperative factors as a cause for concern.. Pediatrics. (2014). 133 e751–7
9. Fleming, S, Thompson, M, Stevens, R, Heneghan, C, Plüddemann, A, Maconochie, I, Tarassenko, L, Mant, D Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies.. Lancet. (2011). 377 1011–8
10. Kheterpal, S Clinical research using an information system: The multicenter perioperative outcomes group.. Anesthesiol Clin. (2011). 29 377–88
11. American Society of Anesthesiologists: American Society of Anesthesiologists - Standards & Guidelines at: http://www.asahq.org/~/media/Sites/ASAHQ/Files/Public/Resources/standards-guidelines/asa-physical-status-classification-system.pdf. Accessed July 22, 2016

12. Borghi, E, de Onis, M, Garza, C, Van den Broeck, J, Frongillo, EA, Grummer-Strawn, L, Van Buuren, S, Pan, H, Molinari, L, Martorell, R, Onyango, AW, Martines, JC WHO Multicentre Growth Reference Study Group, Construction of the World Health Organization child growth standards: Selection of methods for attained growth curves.. Stat Med. (2006). 25 247–65
13. Rigby, RA, Stasinopoulos, DM Automatic smoothing parameter selection in GAMLSS with an application to centile estimation.. Stat Methods Med Res. (2014). 23 318–32
14. van Buuren, S, Fredriks, M Worm plot: A simple diagnostic device for modelling growth reference curves.. Stat Med. (2001). 20 1259–77
15. van Buuren, S, Hayes, DJ, Stasinopoulos, DM, Rigby, RA, ter Kuile, FO, Terlouw, DJ Estimating regional centile curves from mixed data sources and countries.. Stat Med. (2009). 28 2891–911
16. Bijker, JB, van Klei, WA, Kappen, TH, van Wolfswinkel, L, Moons, KG, Kalkman, CJ Incidence of intraoperative hypotension as a function of the chosen definition: Literature definitions applied to a retrospective cohort using automated data collection.. Anesthesiology. (2007). 107 213–20
17. Nafiu, OO, Voepel-Lewis, T, Morris, M, Chimbira, WT, Malviya, S, Reynolds, PI, Tremper, KK How do pediatric anesthesiologists define intraoperative hypotension?. Paediatr Anaesth. (2009). 19 1048–53
18. Bonafide, CP, Brady, PW, Keren, R, Conway, PH, Marsolo, K, Daymont, C Development of heart and respiratory rate percentile curves for hospitalized children. Pediatrics. (2013). 131 e1150–7
19. National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents, The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents.. Pediatrics. (2004). 114 555–76
20. Butani, L, Morgenstern, BZ Are pitfalls of oxcillometric blood pressure measurements preventable in children?. Pediatr Nephrol. (2003). 18 313–8
21. Kaufmann, MA, Pargger, H, Drop, LJ Oscillometric blood pressure measurements by different devices are not interchangeable.. Anesth Analg. (1996). 82 377–81
22. van Buuren, S When to use flexible imputation, Flexible Imputation of Missing Data. (2012). Boca Raton Chapman and Hall/CRC Interdisciplinary Statistics 48–9

Supplemental material

Table 1 Systolic non-invasive blood pressure during preparation phase in relation to age

Preparation	Male					Female				
Phase	Systolic non-invasive blood pressure (mm Hg)					Systolic non-invasive blood pressure (mm Hg)				
Age	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
0	25	35	48	68	101	27	36	46	59	81
1 month	37	49	62	79	104	38	49	60	74	95
2 months	42	53	66	82	106	42	53	65	79	99
3 months	45	57	69	85	107	45	56	68	82	102
4 months	47	59	71	86	108	48	59	71	85	104
5 months	49	61	73	88	108	50	61	73	87	106
6 months	51	63	75	89	109	51	63	74	88	107
9 months	54	66	78	91	110	55	66	78	92	110
1 year	57	68	80	93	111	57	69	80	94	111
2 years	61	73	84	96	112	62	73	84	97	114
3 years	64	75	86	98	115	64	75	86	99	116
4 years	65	77	88	100	117	65	77	88	101	118
5 years	66	78	89	102	120	66	78	90	103	120
6 years	67	79	90	104	122	67	79	91	104	122
8 years	69	80	92	106	125	69	80	92	107	126
10 years	70	82	94	108	128	70	81	93	108	128
12 years	71	83	95	110	132	70	82	94	109	131
14 years	72	84	97	112	136	71	82	95	111	134
16 years	73	85	98	115	141	72	83	96	113	139
18 years	75	87	101	119	148	74	85	98	116	147

Table 2 Mean non-invasive blood pressure during preparation phase in relation to age

Preparation Phase	Male Mean non-invasive blood pressure (mm Hg)					Female Mean non-invasive blood pressure (mm Hg)				
Age	**-2SD**	**-1SD**	**0SD**	**+1SD**	**+2SD**	**-2SD**	**-1SD**	**0SD**	**+1SD**	**+2SD**
0	17	24	33	46	71	18	24	31	42	59
1 month	23	31	40	52	72	24	31	39	50	66
2 months	26	33	42	54	72	26	34	42	53	69
3 months	28	35	44	55	73	28	35	44	55	71
4 months	29	37	45	56	73	29	37	46	57	72
5 months	30	38	46	57	74	30	38	47	58	73
6 months	31	39	47	58	74	31	39	48	59	74
9 months	33	41	49	60	75	33	41	50	61	75
1 year	34	42	50	61	75	35	43	52	62	76
2 years	37	45	53	63	77	37	45	54	65	78
3 years	38	46	55	65	78	38	47	55	66	79
4 years	39	47	56	66	80	39	47	56	67	80
5 years	40	48	57	68	82	40	48	57	68	82
6 years	40	48	58	69	84	41	49	58	70	84
8 years	41	50	60	72	87	42	50	60	72	87
10 years	42	51	61	74	90	43	51	61	73	89
12 years	43	52	62	75	92	44	52	62	75	92
14 years	45	53	63	76	95	44	53	63	76	96
16 years	46	55	65	78	99	46	54	65	79	102
18 years	47	56	66	81	104	47	56	67	82	111

Table 3 Diastolic non-invasive blood pressure during preparation phase in relation to age

Preparation Phase	Male Diastole non-invasive blood pressure (mm Hg)					Female Diastole non-invasive blood pressure (mm Hg)				
Age	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
0	12	17	24	36	60	13	17	23	33	51
1 month	16	22	29	40	59	17	22	29	39	55
2 months	18	24	31	41	59	18	24	31	41	56
3 months	19	25	32	42	59	19	25	33	42	57
4 months	19	26	33	43	59	20	26	34	44	58
5 months	20	26	34	44	59	21	27	35	44	59
6 months	21	27	35	44	59	21	28	35	45	59
9 months	22	28	36	46	60	23	29	37	47	60
1 year	23	29	37	46	60	23	30	38	48	61
1½ year	24	31	38	47	60	24	32	39	49	62
2 years	24	31	39	48	61	25	32	40	49	62
3 years	25	32	40	50	63	25	33	41	50	63
4 years	25	33	41	51	64	25	33	42	51	64
5 years	25	33	42	52	66	25	34	43	52	65
6 years	25	34	43	53	67	26	35	43	53	67
8 years	26	35	44	55	70	26	35	44	55	70
10 years	26	36	45	57	72	27	36	45	56	72
12 years	28	37	46	57	74	28	37	46	57	74
14 years	29	38	47	58	76	30	38	48	59	78
16 years	31	39	48	60	78	31	40	50	62	83
18 years	32	41	50	62	82	34	42	52	66	91

Table 4 Systolic non-invasive blood pressure during surgical phase in relation to age

Surgical Phase	Male					Female				
	Systolic non-invasive blood pressure (mm Hg)					Systolic non-invasive blood pressure (mm Hg)				
Age	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
0	31	41	54	72	103	31	39	50	63	83
1 month	40	51	63	79	104	40	50	62	75	95
2 months	44	55	66	81	104	44	54	66	80	99
3 months	47	57	69	83	105	47	57	69	82	102
4 months	49	59	71	85	106	49	59	71	85	104
5 months	50	61	72	86	106	50	61	73	87	106
6 months	52	62	73	87	107	52	63	74	88	107
9 months	55	65	76	89	108	55	66	78	91	110
1 year	57	67	78	91	109	57	68	80	93	112
2 years	62	72	83	96	113	62	73	85	98	116
3 years	65	75	86	99	117	65	75	87	100	118
4 years	66	77	89	102	120	66	77	89	102	120
5 years	68	78	90	104	123	67	78	90	104	122
6 years	68	79	91	106	124	68	79	91	105	124
8 years	69	81	93	107	127	69	80	93	108	127
10 years	70	82	94	109	128	70	81	94	109	129
12 years	71	83	95	111	130	71	82	94	109	129
14 years	72	84	97	112	133	72	83	95	109	129
16 years	74	86	99	114	135	73	84	96	110	130
18 years	76	88	101	117	137	75	86	97	111	131

Table 5 Mean non-invasive blood pressure during surgical phase in relation to age

Surgical Phase	Male					Female				
	Mean non-invasive blood pressure (mm Hg)					Mean non-invasive blood pressure (mm Hg)				
Age	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
0	21	28	37	49	71	21	27	34	45	62
1 month	26	33	41	51	70	26	32	40	51	67
2 months	27	34	42	53	70	27	34	42	53	69
3 months	29	35	43	53	70	29	36	44	54	71
4 months	30	36	44	54	70	30	37	45	55	72
5 months	31	37	45	55	70	31	38	46	56	73
6 months	31	38	45	55	70	31	39	47	57	73
9 months	33	39	47	56	71	33	40	49	59	75
1 year	34	40	48	57	72	34	42	50	60	76
2 years	36	43	51	61	75	37	44	52	63	78
3 years	38	45	53	63	78	38	45	54	64	79
4 years	38	46	54	65	80	39	46	55	66	81
5 years	39	46	55	67	82	39	47	56	67	82
6 years	39	47	56	68	83	40	48	57	68	84
8 years	40	48	58	70	86	41	49	59	71	86
10 years	42	50	60	72	88	42	50	60	72	88
12 years	43	51	61	74	89	43	51	61	74	89
14 years	44	52	63	76	91	44	52	62	75	90
16 years	45	54	64	77	92	45	53	63	76	92
18 years	47	55	66	79	94	47	55	65	78	95

Table 6 Diastolic non-invasive blood pressure during surgical phase in relation to age

Surgical Phase	Male					Female				
	Diastole non-invasive blood pressure (mm Hg)					Diastole non-invasive blood pressure (mm Hg)				
Age	-2SD	1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	1SD	+2SD
0	15	20	27	38	58	15	19	25	35	53
1 month	17	23	30	39	56	18	23	29	38	56
2 months	18	24	31	40	55	19	24	31	40	57
3 months	19	25	31	40	55	19	25	32	41	57
4 months	19	25	32	41	55	20	26	32	42	58
5 months	20	26	32	41	55	20	26	33	42	58
6 months	20	26	33	41	55	21	27	34	43	59
9 months	21	27	34	42	56	22	28	35	44	59
1 year	21	28	34	43	56	22	29	36	45	60
2 years	22	29	36	45	58	23	30	38	47	61
3 years	23	30	38	47	61	24	31	39	48	62
4 years	23	31	39	49	63	24	32	40	49	64
5 years	23	31	40	50	64	24	32	40	50	65
6 years	23	32	41	51	66	24	33	41	51	66
8 years	24	33	42	53	68	25	34	43	53	68
10 years	25	34	43	54	69	26	35	44	54	69
12 years	26	35	45	55	70	27	36	45	55	70
14 years	28	37	46	56	71	29	37	46	57	72
16 years	30	38	47	57	72	31	39	48	59	74
18 years	32	40	49	59	73	33	41	50	61	78

Table 7 Systolic non-invasive blood pressure during preparation phase in relation to weight

Preparation	Male					Female				
Phase	Systolic non-invasive blood pressure (mm Hg)					Systolic non-invasive blood pressure (mm Hg)				
Weight (kg)	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
2	28	39	50	64	83	30	39	50	62	78
2.5	31	42	54	68	87	33	43	54	67	83
3	34	46	58	71	90	36	47	58	71	87
3.5	37	49	60	74	93	39	50	61	74	91
4	39	51	63	77	96	42	52	64	77	94
4.5	42	54	65	79	98	44	55	66	79	97
5	44	56	67	81	100	46	57	68	82	99
6	47	59	71	84	103	50	61	72	85	103
7	50	62	74	87	105	52	64	75	88	106
8	53	65	76	89	107	55	66	78	91	108
9	55	67	78	91	109	57	68	80	93	111
10	57	69	80	93	111	59	70	81	95	112
12	60	71	83	96	113	61	73	84	97	115
14	62	74	85	98	115	63	75	86	99	117
16	64	76	87	99	117	65	76	88	101	119
18	65	77	88	101	118	66	77	89	102	120
20	67	78	89	102	120	67	78	90	103	121
25	69	80	91	105	123	69	80	91	105	124
30	70	81	93	106	125	69	81	92	106	126
35	70	82	94	108	127	70	81	93	108	128
40	71	82	94	109	130	70	82	94	108	130
50	71	83	96	111	133	71	82	94	110	132
60	72	84	97	113	137	72	83	96	111	135
70	74	86	98	115	140	73	84	97	113	139
80	75	87	100	117	144	75	86	99	116	142
90	76	88	102	119	147					
100	78	90	103	121	150					

Table 8 Mean non-invasive blood pressure during preparation phase in relation to weight

Preparation	Male					Female				
Phase	Mean non-invasive blood pressure (mm Hg)					Mean non-invasive blood pressure (mm Hg)				
Weight (kg)	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
2	19	25	32	42	57	19	26	33	43	56
2.5	21	27	35	45	60	21	28	36	46	59
3	22	29	37	47	62	23	30	38	48	61
3.5	24	31	39	49	64	25	32	40	50	64
4	25	32	40	50	65	26	33	42	52	65
4.5	26	34	42	52	67	27	35	43	53	67
5	27	35	43	53	68	28	36	44	55	69
6	29	37	45	55	70	30	38	47	57	71
7	31	38	47	57	71	32	40	49	59	73
8	32	40	48	59	73	33	41	50	61	74
9	33	41	50	60	74	35	43	51	62	76
10	34	42	51	61	75	36	44	52	63	77
12	36	44	53	63	77	37	45	54	65	79
14	38	46	54	65	79	38	47	55	66	80
16	39	47	56	66	80	39	48	57	67	81
18	40	48	57	68	82	40	48	57	68	83
20	40	49	58	69	83	41	49	58	69	84
25	42	50	59	71	86	42	50	60	71	86
30	42	51	61	72	88	43	51	61	73	89
35	43	52	62	74	90	44	52	62	74	91
40	44	52	62	75	91	44	53	63	75	93
50	45	54	64	76	94	45	54	64	76	95
60	45	54	65	78	96	45	54	64	77	98
70	46	55	65	78	98	46	55	65	79	100
80	47	55	66	79	100	47	55	66	80	103
90	47	56	66	80	101					
100	48	56	66	80	103					

Table 9 Diastolic non-invasive blood pressure during preparation phase in relation to weight

Preparation	Male					Female				
Phase	Diastolic non-invasive blood pressure (mm Hg)					Diastolic non-invasive blood pressure (mm Hg)				
Weight (kg)	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
2	13	18	23	32	46	14	18	24	32	45
2.5	15	19	25	34	48	15	20	26	34	47
3	16	21	27	35	50	16	21	28	36	49
3.5	16	22	28	37	51	17	23	29	38	51
4	17	23	29	38	52	18	24	30	39	52
4.5	18	24	30	39	53	19	25	32	41	54
5	18	24	31	40	54	19	25	33	42	55
6	20	26	33	42	56	21	27	34	44	57
7	20	27	34	44	57	22	28	36	45	58
8	21	28	35	45	58	22	29	37	46	60
9	22	29	36	46	59	23	30	38	47	61
10	22	30	37	47	60	23	31	39	48	62
12	23	31	39	48	62	24	32	40	50	63
14	24	32	40	50	63	25	33	41	51	64
16	25	33	41	51	64	25	33	42	51	65
18	25	33	42	52	65	26	34	42	52	66
20	25	34	42	53	66	26	34	43	53	67
25	26	35	44	54	69	27	35	44	54	69
30	27	36	45	55	70	27	36	45	56	71
35	27	36	45	56	72	28	37	46	57	73
40	28	37	46	57	73	29	38	47	58	74
50	29	38	47	58	75	30	39	48	59	77
60	30	39	48	59	76	31	39	49	61	79
70	31	39	48	59	77	32	40	49	62	82
80	32	40	48	60	78	33	41	50	63	84
90	32	40	48	60	79					
100	33	40	48	60	81					

Table 10 Systolic non-invasive blood pressure during surgical phase in relation to weight

Surgical Phase	Male					Female				
	Systolic non-invasive blood pressure (mm Hg)					Systolic non-invasive blood pressure (mm Hg)				
Weight (kg)	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
2	33	43	53	66	85	32	42	51	62	78
2.5	36	46	56	69	88	36	45	55	67	83
3	38	48	59	72	91	39	48	58	70	87
3.5	40	51	61	74	93	41	51	61	74	91
4	42	53	63	76	95	43	54	64	76	94
4.5	44	55	65	78	97	45	56	66	79	96
5	46	56	67	80	99	47	58	69	81	99
6	49	59	70	83	102	50	61	72	85	103
7	51	62	73	86	104	53	64	75	88	106
8	54	64	75	88	106	55	66	77	91	109
9	56	66	77	90	108	57	68	80	93	111
10	57	68	79	92	110	59	70	81	95	113
12	60	71	82	95	113	62	73	84	98	116
14	63	73	85	98	116	64	75	86	100	118
16	65	75	87	100	118	65	76	88	101	120
18	66	77	88	102	120	67	77	89	103	121
20	68	78	90	103	121	68	78	90	104	122
25	69	80	92	105	124	69	80	91	105	124
30	70	81	93	107	125	70	81	92	106	125
35	71	82	94	108	127	71	81	93	107	126
40	71	82	94	109	128	71	82	93	107	126
50	72	83	96	111	130	72	83	94	108	128
60	73	85	98	113	132	73	84	96	110	129
70	75	87	100	115	135	75	86	98	112	132
80	76	88	102	117	137	77	88	100	114	135
90	77	90	103	119	139					
100	78	91	105	121	141					

Table 11 Diastolic non-invasive blood pressure during surgical phase in relation to weight

Surgical	Male					Female				
Phase	Mean non-invasive blood pressure (mm Hg)					Mean non-invasive blood pressure (mm Hg)				
Weight (kg)	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
2	23	29	36	44	60	22	28	35	44	58
2.5	24	30	37	46	61	24	30	37	46	61
3	25	31	38	47	63	25	31	39	48	63
3.5	26	32	39	49	64	26	33	40	50	65
4	27	33	40	50	65	27	34	42	51	66
4.5	28	34	41	51	66	28	35	43	52	68
5	29	35	42	52	66	29	36	44	54	69
6	30	36	44	53	68	31	38	46	56	71
7	31	38	45	55	69	32	39	47	57	73
8	32	39	46	56	71	33	40	49	59	74
9	33	40	48	58	72	34	41	50	60	75
10	34	41	49	59	73	35	42	51	61	77
12	36	43	51	61	75	36	44	52	63	78
14	37	44	52	63	77	38	45	54	64	80
16	38	45	54	64	79	38	46	55	66	81
18	39	46	55	66	80	39	47	56	67	82
20	40	47	56	67	82	40	48	57	68	83
25	41	49	58	69	84	41	49	58	70	85
30	42	50	59	71	86	42	50	59	71	86
35	42	51	60	72	88	43	51	60	72	88
40	43	51	61	74	89	44	52	61	73	88
50	44	53	63	75	90	45	53	62	75	90
60	45	54	64	77	92	45	53	63	76	91
70	46	54	65	78	93	46	54	65	77	93
80	46	55	66	79	93	47	55	66	79	95
90	47	56	66	79	94					
100	47	56	67	80	95					

Table 12 Diastolic non-invasive blood pressure during surgical phase in relation to weight

Surgical Phase	Male					Female				
	Diastolic non-invasive blood pressure (mm Hg)					Diastolic non-invasive blood pressure (mm Hg)				
Weight (kg)	-2SD	-1SD	0SD	+1SD	+2SD	-2SD	-1SD	0SD	+1SD	+2SD
2	16	21	26	33	46	16	20	25	33	47
2.5	17	21	27	34	47	17	21	27	34	49
3	17	22	28	35	48	17	22	28	36	50
3.5	18	23	29	36	49	18	23	29	37	52
4	18	23	29	37	50	19	24	30	38	53
4.5	18	24	30	38	51	19	25	31	39	54
5	19	24	30	38	52	19	25	32	40	55
6	19	25	31	40	53	20	26	33	42	57
7	20	26	32	41	54	21	27	34	43	58
8	20	27	33	42	55	21	28	35	44	59
9	21	27	34	43	56	22	29	36	45	60
10	21	28	35	44	57	22	29	36	46	61
12	22	29	36	45	59	23	30	38	47	62
14	22	30	37	47	60	23	31	39	48	63
16	23	30	38	48	62	24	31	39	49	64
18	23	31	39	49	63	24	32	40	50	64
20	23	32	40	50	64	24	32	41	50	65
25	24	33	42	52	66	25	34	42	52	66
30	25	34	43	53	68	26	34	43	53	67
35	26	35	44	54	69	27	35	44	54	68
40	26	35	44	55	69	28	36	45	55	69
50	28	37	46	56	70	29	37	46	57	71
60	29	38	47	57	71	30	39	47	58	73
70	30	39	47	57	71	31	40	49	59	74
80	31	39	48	58	72	32	40	49	60	76
90	32	40	48	58	72					
100	32	40	48	58	73					

3

Patient and anesthesia characteristics of children with low pre-incision blood pressure, a retrospective observational study

Wietze Pasma
Linda M. Peelen
Stefanie van den Broek
Stef van Buuren
Wilton A. van Klei
Jurgen C. de Graaff

Published in Acta Anaesthesiologica Scandinavica 2020 Apr;64(4):472-480

Abstract

Background

Intraoperative blood pressure has been suggested as a key factor for safe pediatric anesthesia. However, there is not much insight into factors that discriminate between children with low and normal pre-incision blood pressure. Our aim was to explore if children who have a low blood pressure during anesthesia are different than those with normal blood pressure. The focus of the present study was on the pre-incision period.

Methods

This retrospective study included pediatric patients undergoing anesthesia for non-cardiac surgery at a tertiary pediatric university hospital, between 2012 and 2016. We analyzed the association between pre-incision blood pressure and patient- and anesthesia characteristics, comparing low with normal pre-incision blood pressure. This association was further explored with a multivariable linear regression.

Results

In total, 20 962 anesthetic cases were included. Pre-incision blood pressure was associated with age (beta -0.04 SD per year), gender (female -0.11), previous surgery (-0.15), preoperative blood pressure (+0.01 per mmHg), epilepsy (0.12), bronchial hyperactivity (-0.18), emergency surgery (0.10), locoregional technique (-0.48), artificial airway device (supgraglottic airway device instead of tube 0.07 and sevoflurane concentration (0.03 per sevoflurane %).

Conclusions

Children with low pre-incision blood pressure do not differ on clinically relevant factors from children with normal blood pressure. Although the present explorative study shows that pre-incision blood pressure is partly dependent on patient characteristics and partly dependent on anesthetic technique, other unmeasured variables might play a more important role.

Introduction

During anesthesia vital signs such as blood pressure are monitored according to standards and guidelines.[1] In 2016 reference curves for age-appropriate blood pressure measurements under anesthesia were developed, using data from over 100,000 children across 11 centers.[2] These curves show the relation between age, weight or height and blood pressure during anesthesia, and allow us to compare these with actual blood pressure measurements during surgical care. The references were developed for a relatively stable period, most likely not influenced by anesthetic and surgical factors (e.g. post-induction dip in blood pressure and stress reaction on incision). In the next step, it has to be elucidated which children fall below these references or, in other words, which patients are at outliers of the reference values?

In adults, there is evidence that older age, higher ASA-physical status and co-existing conditions such as hypertension, diabetes mellitus and myocardial infarction are associated with intraoperative hypotension. Additionally, in adults researchers reported that intraoperative low blood pressure is associated with organ injury and adverse outcomes such as prolonged postoperative stay and death. On the other hand in the pediatric population this evidence is not present, although blood pressure has been suggested as a key factor for safe pediatric anesthesia.[3–8] In contrast to research in adults, only a few studies are available in children, where fasting status, ASA physical status, preoperative hypotension, intravenous induction, propofol dosage and body mass index were found to be associated with intraoperative hypotension in children.[9–11]

The aim of this study was to explore if children who have a low blood pressure during anesthesia are different than those with normal blood pressure. The focus of the present study was on the pre-incision period. We hypothesized that several preexisting patient characteristics are associated with pre-incision blood pressure and that differences in the management of anesthesia induction, such as differences in medication dosage are associated with pre-incision blood pressure.

Methods

This retrospective cohort study included all non-cardiac pediatric anesthetic procedures performed at a specialized tertiary referral university hospital (Wilhelmina Children's Hospital Utrecht, the Netherlands), from January 1st 2012

until December 31st 2016. Similar to the previous study in which the reference curves were developed[2], we excluded cardiac procedures or when the surgical specialty was missing. Also, at least two blood pressure measurements had to be available before incision. If the time of incision was not available, the procedure was excluded. All data were retrospectively collected from the Anesthesia Information Management System (AIMS, Anstat, version 2.0.4, 2015, Carepoint, Ede, The Netherlands) and Electronic Health Record (HiX, Chipsoft, Amsterdam, The Netherlands). The IRB waived the need for informed consent under the Dutch Data Protection Act (METC number 16/235). We de-identified the data before analysis.

Blood pressure

We based low pre-incision blood pressure on non-invasive mean arterial blood pressure measurements, as this parameter, rather than systolic or diastolic blood pressure, is the key parameter in the local protocol for intraoperative blood pressure control. Non-invasive blood pressure is measured according to protocol at least every 5 minutes by oscillometry and stored in the AIMS database. For our definition of low pre-incision blood pressure we collected measurements within 20 minutes before the start of the procedure (marked by an event registration of start incision) and calculated the mean of the last three of these measurements, which was the same method as was used in the development of the previously published references. Before the collection of measurements, we removed measurements that were defined as artifacts, i.e., when the diastolic pressure was lower than 3 mmHg, when the systolic pressure was equal or higher than 250 mmHg, when the pulse pressure (systolic pressure minus diastolic pressure) was lower than or equal to 5 mmHg or when one of the systolic, diastolic or mean arterial pressure values was missing.[2] Subsequently, using the reference curves, we calculated a standardized pre-incision blood pressure (Z score) given the patient's height and gender using the relevant reference curve for mean arterial blood pressure values in the pre-incision period.[2] We collected height values within a clinically relevant time period before surgery, whereby this period depended on patient age (see supplemental data S1). If no height measurement was available within this period, we considered height as missing data. Finally, we defined low pre-incision blood pressure as a standardized blood pressure value (Z) lower than -2 standard deviations (SD) (ranging from 19 to 48 mmHg, for height 45 to 200 cm). We considered a standardized blood pressure between -2 SD and 2 SD (55 to 105 mmHg, for height 45 to 200 cm) as normal, and standardized blood pressure above

2 SD as high.[2] We purposefully do not define hypotension in this study, which would imply that the blood pressure below a threshold is too low and harmful. Since there is no clear consensus on a hypotension definition for pre-incision blood pressure in children under anesthesia, the choice of cut-off value in this paper was arbitrary.

Patient and anesthesia characteristics

Characteristics were divided into patient and anesthesia related characteristics. Patient characteristics collected for this study were gender, age, preoperative height, preoperative weight, ASA physical status, preoperative blood pressure (in mmHg), time of the start of the procedure(morning (8AM to 12AM), afternoon (12AM to 5 PM), evening till midnight (5PM to 12 PM) and after midnight (12PM to 8 AM), previous surgery and pre-operative comorbidities. These factors have been studied before and were associated with blood pressure in children or in adults.[3,9,10]

We collected comorbidity data from preoperative evaluation charts, where we focused on comorbidities we assumed likely related to intraoperative blood pressure. When preoperative evaluation was performed but information on comorbidities was (partly) missing, we assumed that these comorbidities were not present. Procedure characteristics which we considered were: surgical specialty, priority status, anesthetic technique performed, medication use during pre-incision phase (propofol, atracurium and sufentanil) and the inspired sevoflurane concentration.[3,10,12–15] The start of the pre-incision period during which medication data was collected, was defined by the first of the three bloodpressure measurements that were used to calculate the pre-incision blood pressure. The end of this period was equal to the start of the surgical procedure. From this period, the median of inspired sevoflurane concentration was used.

Statistical analyses

In the first part of the analyses we described the characteristics of the patients and the procedures in which a low pre-incision blood pressure occurred and compared these characteristics to children with a normal pre-incision blood pressure. Hence, we excluded cases with a relatively high blood pressure (>2 SD) for this part of the analysis. For continuous data, the median and interquartile range is presented. Because of the large sample size we assumed the variance to be normally

distributed, and compared the groups using a *t*-test. In case of categorical and dichotomous variables, the data are presented as counts and percentages and groups are compared with a chi-squared test.

In the second part of the analyses, we assessed the association between patient and anesthesia characteristics versus blood pressure using the calculated Z-scores as a continuous outcome variable using multivariable linear regression analysis. Cases with a high blood pressure (Z > 2SD) were included in these analyses. We included the same patient characteristics – excluding height, weight and ASA physical status, because of expected collinearity – and anesthesia characteristics into a linear regression model with standardized pre-incision blood pressure as outcome. This first and second part of the analysis were defined before prior to obtaining the data. The study was designed and reported according to the STROBE guidelines.

As a post-hoc analysis, to investigate whether potential risk factors were different for infants and older children, we fitted the same model in children younger than 12 months and children older than 12 months separately.

Missing data rarely occur completely at random and conducting complete case analysis typically leads to biased effect estimates.[16,17] Therefore we used multiple imputation using the *mice* package.[18] We imputed twenty complete datasets, in which we used passive imputation to impute Z values for pre-incision blood pressure. We pooled effect estimates and test statistics in individual imputation sets using Rubin's rules.[16] Results presented throughout the manuscript are based on these imputed data.

We extracted and processed the data from our local enterprise data warehouse, using SAS software (Version 9.4, Copyright © 2013 SAS Institute Inc). We further processed and analyzed the de-identified data in R (R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org, version 3.3.2 (2016-10-31)). We considered a p-value < 0.05 statistically significant throughout the analyses.

Results

Cohort selection

Within the study period, we identified 31 984 pediatric anesthesia procedures. We excluded 2 877 (9%) cases as these were cardiac procedures or the surgical specialty was missing. An additional 8 145 (25%) cases were excluded because the number of measurements was too low (less than 2). This resulted in 20 962 included procedures (Figure 1). Distribution of gender, age, ASA physical status and surgical specialty of all procedures and those included in our analysis are provided in table 1.

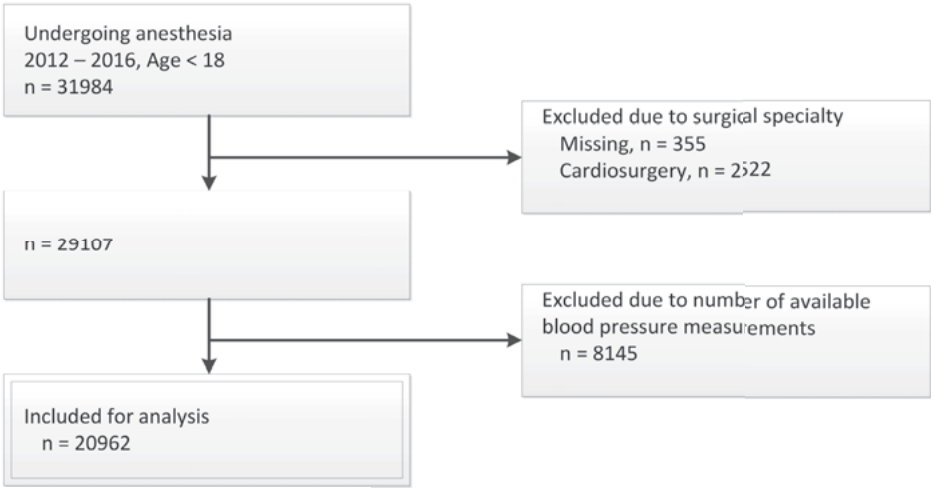


Figure 1: Flowchart of included anesthesia procedures for analysis.
Double framed boxes indicate the included procedures that were used for analysis. 20 962 procedures were included for the multivariable regression analyses.
* Measurements of height were considered expired, when it was not measured within a clinical relevant time before surgery (supplemental material)

Table 1: Baseline characteristics of included anesthetic procedures (n = 20 962)

Parameter		N	(%)
Number of cases		20962	
Age	Neonates (0 - 1 month)	516	(2.5%)
	Infant (1 month - 1 year)	3342	(15.9%)
	1 to 4 years	4776	(22.8%)
	4 to 8 years	4132	(19.7%)
	8 to 18 years	8196	(39.1%)
Gender	Male	8440	(40.3%)
	Female	12522	(59.7%)
ASA physical status	1	10213	(48.7%)
	2	6644	(31.7%)
	>2	1541	(7.4%)
	Unknown	2564	(12.2%)
Surgical Specialty	Pediatric Surgery	4509	(21.5%)
	Maxillofacial	828	(4.0%)
	Neurosurgery	872	(4.2%)
	Ophthalmology	1549	(7.4%)
	Otolaryngologic surgery	3461	(16.5%)
	Pediatric intervention	3916	(18.7%)
	Reconstructive surgery	1473	(7.0%)
	Urologic surgery	4354	(20.8%)

Categorical data is presented as number of procedures and percentage.

We selected three blood pressure measurements within 20 minutes before incision, which we used to calculate the standardized blood pressure. The median period in which these measurements were selected, was 10 minutes (IQR 7 – 12). The within patient variation of these blood pressure measurements was low. The median of the standard deviation per patient was 4.2 mmHg (IQR 2.1 – 7.8 mmHg) and the median of the range (maximum minus minimum) was 8 (IQR 4 – 14).

Low versus normal blood pressure

In total 6.1% (n = 1 259) of the procedures patients had a low pre-incision blood pressure. For this part of the analysis 477 cases with high pre-incision blood

pressure were excluded. The comparison of patient and anesthesia characteristics showed that children with low blood pressure were older, more often female, weighed more, were longer in height, had a higher pre-operative blood pressure, had more often previous surgery, were more often operated after office hours (after 5PM), had more often movement disorders and had more often kidney disorders (Table 2a). In addition these children underwent procedures of different surgical disciplines. Emergency surgery was less common in children with a low pre-incision blood pressure. The combination of general and a locoregional anesthesia and a tube (instead of a supraglottic airway) was more often used in the low blood pressure group. Dosage of pre-incision medication was similar in both groups, and concentration of sevoflurane was lower in the low blood pressure group (Table 2b).

Table 2a: Comparison of patient characteristics between anesthesia procedures with low and normal pre-incision blood pressure

Parameter		**Low blood pressure**		**Normal blood pressure**		**p-value**
Group size		1259		19226		
Age (years)		8.1	(4.8 - 12.7)	5.5	(1.4 - 11.3)	**
Patient gender	Male	716	(56.9 %)	11533	(60 %)	*
Patient weight (kg)		26	(17.6 - 43)	19.6	(10.7 - 37)	**
Patient height (cm)		129	(105 - 153)	110	(78.3 - 145)	**
ASA physical status	1	731	(58.1 %)	10583	(55 %)	
	2	413	(32.8 %)	6928	(36 %)	
	>2	115	(9.1 %)	1715	(8.9 %)	
Pre-operative NIBP		77	(71 - 84)	78	(72 - 85)	
Time of surgery	After midnight	15	(1.2 %)	202	(1.1 %)	*
	Morning	389	(30.9 %)	6329	(32.9 %)	
	Afternoon	814	(64.6 %)	11725	(61.0 %)	
	Evening till midnight	42	(3.3 %)	970	(5.0 %)	
Patient had previous surgery		917	(72.9 %)	12097	(62.9 %)	**
Bleeding disorders		24	(1.9 %)	424	(2.2 %)	
Cardiac history		134	(10.7 %)	1734	(9 %)	

Table 2a *(continued)*: Comparison of patient characteristics between anesthesia procedures with low and normal pre-incision blood pressure

Parameter	Low blood pressure	Normal blood pressure	p-value
Coagulation disorders	29 (2.3 %)	499 (2.6 %)	
Bronchial hyperreactivity	18 (1.4 %)	267 (1.4 %)	
Movement disorder	158 (12.5 %)	1871 (9.7 %)	*
Apnea	23 (1.8 %)	274 (1.4 %)	
Epilepsy	45 (3.6 %)	831 (4.3 %)	
Kidney disorders	106 (8.4 %)	1279 (6.7 %)	*
Liver disorders	16 (1.3 %)	218 (1.1 %)	
Lung disorders	138 (11.0 %)	1958 (10.2 %)	
Recurrent airway disorders	21 (1.7 %)	386 (2.0 %)	

Low blood pressure is defined as values below -2SD and normal blood pressure as values between -2SD and +2SD. Pre-incision reference values, corrected for height and gender were used. Due to pooling and rounding of results, these numbers might not add up to group totals. Continuous data is presented as median and interquartile range, categorical data is presented as number of procedures and percentage. * p-value < 0.05 ** p-value < 0.001

Table 2b: Comparison of procedure characteristics between anesthesia procedures with low and normal pre-incision blood pressure

Parameter		Low blood pressure		Normal blood pressure		p-value
Group size		1259		19226		
Surgical discipline	Pediatric surgery	308	(24.5 %)	4143	(21.5 %)	**
	Maxillofacial	52	(4.2 %)	772	(4.0 %)	
	Neurosurgery	40	(3.2 %)	814	(4.2 %)	
	Ophthalmology	61	(4.8 %)	1462	(7.6 %)	
	Otolaryngologic surgery	274	(21.8 %)	3081	(16.0 %)	
	Pediatric intervention	110	(8.7 %)	3579	(18.6 %)	
	Reconstructive surgery	94	(7.5 %)	1371	(7.1 %)	
	Urologic surgery	319	(25.3 %)	4005	(20.8 %)	
Priority status of surgery	Emergency	159	(12.6 %)	3084	(16.0 %)	*
	Planned	1100	(87.4 %)	16142	(84.0 %)	
Locoregional technique used		486	(38.6 %)	4946	(25.7 %)	**
Artificial airway used	Supraglottic airway	577	(45.8 %)	9682	(50.4 %)	*
	Tube	682	(54.2 %)	9544	(49.6 %)	
Inspired sevoflurane (%)		2.7	(2.1 - 3.4)	3.0	(2.3 - 3.8)	**
Propofol (mg/kg) *		0.0	(0.0 - 3.1)	0.0	(0.0 - 2.9)	
Sufentanil (mg/kg) *		0.1	(0.0 - 0.1)	0.1	(0.0 - 0.1)	*
Atracurium (mg/kg) *		0.0	(0.0 - 0.4)	0.0	(0.0 - 0.4)	*

Low blood pressure is defined as values below -2SD and normal blood pressure as values between -2SD and +2SD. Pre-incision reference values, corrected for height and gender were used. Due to pooling and rounding of results, these numbers might not add up to group totals. * If this anesthetic medication was not given, dose of zero mg/kg is included in the analysis. Continuous data is presented as median and interquartile range, categorical data is presented as number of procedures and percentage. * p-value < 0.05 ** p-value < 0.001

Characteristics versus blood pressure

Figure 2 present the results from the multivariable regression analyses using the continuous Z values as an outcome and both patient and anesthesia characteristics as determinants. Older children (beta -0.04 SD per year (95% confidence interval -0.05 - -0.04)) and females (-0.11 (-0.14 - -0.08)) had a lower pre-incision blood pressure, as did children who underwent surgery previously (-0.15 (-0.19 - -0.11)) or those with a lower preoperative blood pressure (0.01 (0.01 – 0.01)). Children with epilepsy (0.12 (0.04 – 0.20)) had a higher pre-incision blood pressure and bronchial hyperreactivity (-0.18 (-0.31 - -0.05)) was associated with a lower pre-incision blood pressure.

The use of a locoregional technique additionally to general anesthesia (-0.48 (-0.53 - -0.44)) was associated with a lower pre-incision blood pressure. The use of supraglottic airway device (laryngeal mask airway) (0.07 (0.00 – 0.14)) was associated with a higher pre-incision blood pressure, compared to the use of an endotracheal tube. The dosage of propofol, atracurium and sufentanil was not associated with pre-incision blood pressure. In contrast a higher concentration of sevoflurane (0.03 per sevoflurane % (0.02 – 0.05)) was associated with an increase in pre-incision blood pressure.

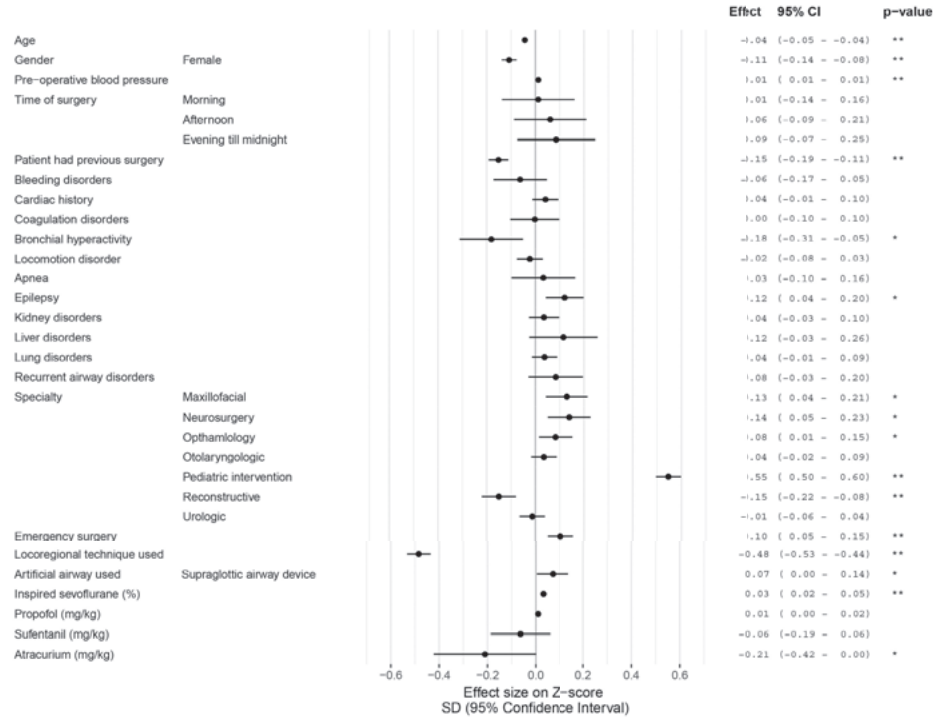


Figure 2: Forest plot of results of linear regression model for association between patient and procedure characteristics vs standardized pre-incision blood pressure (Z score). Effect sizes are in Z score (unit is 1 standard deviation (SD)) with a 95% confidence interval (95% CI) of pre-incision blood pressure. For time of surgery the reference was midnight till morning, for specialty the reference was pediatric surgery and for artificial airway the reference was tube. * p-value <0.05, ** p-value <0.001

Different age groups

As a post-hoc analysis, the model was fitted in infants and older children separately. The model for children under 12 months (table 3a) showed fewer significant associations than the model with the older children (table 3b). Age, pre-operative blood pressure, surgical specialty and use of locoregional technique are associated with pre-incision blood pressure in both groups. Gender, previous surgery, epilepsy, inspired sevoflurane and sufentanil dosage are associated with pre-incision blood pressure in children over 12 months, but not in infants; whereas lung disorders and choice of artificial airway are associated with pre-incision blood pressure in infants, but not in older children.

Table 3a: Multivariable linear regression analysis of factors associated with standardized pre-incision blood pressure in infants, defined as children younger than 12 months of age. (n = 3858)

Parameter		**Effect estimate (95% CI)**		**p-value**
Age (years)		0.25	(0.11 - 0.38)	**
Female		-0.03	(-0.11 - 0.04)	
Pre-operative blood pressure		0.01	(0.01 - 0.02)	**
Time of surgery	Midnight till morning (reference)			
	Morning	-0.13	(-0.45 - 0.18)	
	Afternoon	-0.18	(-0.49 - 0.13)	
	Evening till midnight	-0.07	(-0.41 - 0.26)	
Patient had previous surgery		0.02	(-0.07 - 0.11)	
Bleeding disorders		0.19	(-0.33 - 0.71)	
Cardiac history		0.08	(-0.04 - 0.20)	
Coagulation disorders		-0.05	(-0.39 - 0.29)	
Bronchial hyperreactivity		-0.02	(-0.39 - 0.35)	
Locomotion disorder		-0.02	(-0.25 - 0.21)	
Apnoe		0.09	(-0.32 - 0.49)	
Epilepsy		-0.03	(-0.32 - 0.26)	
Kidney disorders		0.10	(-0.02 - 0.22)	
Liver disorders		0.00	(-0.35 - 0.34)	
Lung disorders		0.19	(0.05 - 0.32)	*
Recurrent airway disorders		-0.09	(-0.33 - 0.14)	

Table 3a *(continued)*: Multivariable linear regression analysis of factors associated with standardized pre-incision blood pressure in infants, defined as children younger than 12 months of age. (n = 3858)

Parameter		Effect estimate (95% CI)		p-value
Surgical specialty	Pediatric surgery (reference)			
	Maxillofacial	-0.45	(-1.15 - 0.26)	
	Neurosurgery	0.00	(-0.17 - 0.16)	
	Opthalmology	-0.10	(-0.33 - 0.14)	
	Otolaryngologic surgery	0.43	(0.29 - 0.57)	**
	Pediatric intervention	0.39	(0.26 - 0.51)	**
	Reconstructive surgery	-0.11	(-0.25 - 0.03)	
	Urologic surgery	0.10	(-0.01 - 0.20)	
Emergency surgery		0.04	(-0.06 - 0.13)	
Locoregional technique used		-0.28	(-0.38 - -0.19)	**
Artificial airway used	Tube (reference)			
	Supraglottic airway device	0.21	(0.12 - 0.31)	**
Inspired sevoflurane (%)		0.02	(0.00 - 0.05)	
Propofol (mg/kg)		-0.02	(-0.04 - 0.00)	
Sufentanil (mg/kg)		0.18	(-0.02 - 0.37)	
Atracurium (mg/kg)		-0.08	(-0.23 - 0.07)	

Effect sizes are presented as betas and should be interpreted as follows: an increase of one unit of the covariate will increase the blood pressure Z-value (standardized pre-incision non-invasive blood pressure) by beta times the SD and 95% confidence intervals (CI). * p-value < 0.05 ** p-value < 0.001

Table 3b: Multivariable linear regression analysis of factors associated with standardized pre-incision blood pressure in children older than 12 months of age. (n = 17104)

Parameter		**Effect estimate (95% CI)**		**p-value**
Age (years)		-0.02	(-0.03 - -0.02)	**
Female		-0.10	(-0.13 - -0.06)	**
Pre-operative blood pressure		0.01	(0.01 - 0.02)	**
Time of surgery	Midnight till morning (reference)			
	Morning	0.04	(-0.12 - 0.20)	
	Afternoon	0.11	(-0.06 - 0.27)	
	Evening till midnight	0.15	(-0.03 - 0.32)	
Patient had previous surgery		-0.08	(-0.13 - -0.04)	**
Bleeding disorders		-0.05	(-0.16 - 0.06)	
Cardiac history		0.01	(-0.05 - 0.07)	
Coagulation disorders		0.01	(-0.09 - 0.11)	
Bronchial hyperreactivity		-0.13	(-0.28 - 0.01)	
Locomotion disorder		0.00	(-0.05 - 0.05)	
Apnoe		0.10	(-0.04 - 0.23)	
Epilepsy		0.16	(0.08 - 0.24)	**
Kidney disorders		-0.06	(-0.13 - 0.01)	
Liver disorders		0.13	(-0.02 - 0.29)	
Lung disorders		0.01	(-0.04 - 0.07)	
Recurrent airway disorders		0.09	(-0.04 - 0.21)	

Table 3b: Multivariable linear regression analysis of factors associated with standardized pre-incision blood pressure in children older than 12 months of age. (n = 17104)

Parameter		Effect estimate (95% CI)		p-value
Surgical specialty	Pediatric surgery (reference)			
	Maxillofacial	0.21	(0.12 - 0.30)	**
	Neurosurgery	0.27	(0.17 - 0.38)	**
	Opthalmology	0.21	(0.14 - 0.28)	**
	Otolaryngologic surgery	0.10	(0.04 - 0.15)	*
	Pediatric intervention	0.65	(0.60 - 0.71)	**
	Reconstructive surgery	-0.08	(-0.16 - 0.00)	*
	Urologic surgery	0.11	(0.05 - 0.18)	**
Emergency surgery		0.06	(0.00 - 0.12)	*
Locoregional technique used		-0.61	(-0.67 - -0.55)	**
Artificial airway used	Tube (reference)			
	Supraglottic airway device	0.07	(-0.03 - 0.17)	
Inspired sevoflurane (%)		0.04	(0.03 - 0.06)	**
Propofol (mg/kg)		0.02	(0.00 - 0.04)	*
Sufentanil (mg/kg)		-0.29	(-0.46 - -0.13)	**
Atracurium (mg/kg)		-0.30	(-0.65 - 0.05)	

Effect sizes are presented as betas and should be interpreted as follows: an increase of one unit of the covariate will increase the blood pressure Z-value (standardized pre-incision non-invasive blood pressure) by beta times the SD and 95% confidence intervals (CI). * p-value < 0.05 ** p-value < 0.001

Discussion

Overall results

We were not able to identify a 'typical' child or procedure prone to have a low pre-incision blood pressure. Instead, the group of children with a low blood pressure was quite heterogeneous. We found several associations with pre-incision blood pressure, of which the association with the use of a locoregional technique was the most profound.

At first glance, patient factors such as age, preoperative blood pressure, bronchial hyperreactivity, epilepsy, kidney disorders, and surgical specialty had the largest influence on pre-incision blood pressure. In comparison, in the anesthetic management, only the use of locoregional anesthesia, the type of airway management and sevoflurane concentration were factors associated with blood pressure. It is likely though that other unmeasured factors are taken in account by the anesthesiologist, which we are unable to collect objectively in retrospective data, for example anesthesia dosing strategy based on clinical experience and intuition. These factors may have a larger influence on blood pressure than those included in this study.

Comparison with previous research

When we compare our findings with previous research, we should first emphasize that the definition of hypotension in previous studies was different. For example Nafiu et al., used reference values based on blood pressure in awake children, instead of reference values in children under anesthesia.[2,10] Others, for example Weber et al, defined low blood pressure as a drop of blood pressure relative to the pre-operative blood pressure of the patients.[19] The factors age and pre-operative blood pressure which we found to be associated with pre-incision blood pressure in this study, have been reported previously[10,19]

If we compare the association of pre-incision blood pressure with patient and anesthesia characteristics between infants and children older than 12 months, we did observe some differences. But we have to take into account that the number of infants was considerably smaller than the overall population (n = 3858; 20% of overall population) and that some of the factors hardly occurred in this age group, such as for example epilepsy. This has lowered the power to detect a potential association.

Strengths and limitations

To our knowledge, only a limited number of studies have tried to characterize children with a low pre-incision blood pressure.[9,11,20] The large sample size and broad inclusion criteria are strengths of the present study. We aimed to facilitate interpretation by using gender and height adjusted references for normalizing mean blood pressure.

This study was designed to evaluate the reference values for blood pressure during anesthesia which we published previously. We aimed to gain more information about those children who are below the normal range. Therefore, we preferred to use the same data collection method to prevent bias, i.e. to focus only on the pre-incision period, using the last three measurements before incision. We did not explore other methodological options, such as analyzing the pattern of all or part of the blood pressure measurements or summarizing blood pressures in a hypotension metric, for example the minimum blood pressure or the area under a threshold. These methods quantify intraoperative hypotension, requiring a choice in method and, in most cases, a hypotension threshold, which is a source of discussion in hypotension research in adults.[21,22] Although hypotension quantification was not the aim of the present study, this could be interesting for future research. The current study can be viewed as an initial step in this direction, because quantifying hypotension using a standardized blood pressure makes more sense than applying the same hypotension definition to neonates, children and young adults.

When interpreting the findings of this study, we also need to acknowledge several limitations. First of all, this study was designed as an retrospective observational explorative study. Hence, the findings of this study cannot directly be generalized to clinical practice at this point but should be considered as hypothesis generating for further research. Secondly, the reference curves we used were developed on a multi-center dataset, containing predominately US centers, including relatively healthy children (ASA physical status 1 and 2).[2] Our study population was different to this population, since we also included higher ASA physical status. Consequently, potential misfit of the reference curves may have biased our findings. We have some indication that this phenomenon existed in the data, which we studied. We expected age not to be associated with blood pressure, because we adjusted blood pressure indirectly for age when we calculated the Z values. Nevertheless,

age remained associated after a multivariate model was fit. We also cannot rule out that other factors were associated with blood pressure, as a consequence of collinearity with age. Thirdly we did not take into account measurement aspects of pre-incision blood pressure such as patient position, blood pressure cuff size used, or manipulation of the patient during measurement. These aspects were not reliably registered in our AIMS data. Not taking measurement aspects into account may lead to artifacts, extra variation or bias in the data.[23]

Conclusion

In conclusion, the population of children with a low pre-incision blood pressure is heterogeneous and therefore we cannot describe a typical pediatric patient prone for low blood pressure during surgery. Although pre-incision blood pressure is associated with choices in anesthesia technique, for example with loco-regional anesthesia technique, we do not think that the data of the present study indicate that current clinical practice should be altered in favor of the intra-operative blood pressure. The data presented is a representation of a safe anesthesia practice, in which low blood pressure can occur and is also managed adequately.

Conflict of interest

None of the authors have a conflict of interest regarding this research project.

References

1. American Society of Anesthesiologists: American Society of Anesthesiologists - Standards & Guidelines at <http://www.asahq.org/quality-and-practice-management/standards-and-guidelines>
2. Graaff JC de, Pasma W, Buuren S van, Duijghuisen JJ, Nafiu OO, Kheterpal S, Klei WA van: Reference Values for Noninvasive Blood Pressure in Children during Anesthesia: A Multicentered Retrospective Observational Cohort Study. Anesthesiology 2016; 125:904–13
3. Reich DL, Hossain S, Krol M, Baez B, Patel P, Bernstein A, Bodian CA: Predictors of hypotension after induction of general anesthesia. Anesth Analg 2005; 101:622–8
4. Bijker JB, Klei WA van, Vergouwe Y, Eleveld DJ, Wolfswinkel L van, Moons KGM, Kalkman CJ: Intraoperative hypotension and 1-year mortality after noncardiac surgery. Anesthesiology 2009; 111:1217–26
5. Waes JAR van, Klei WA van, Wijeysundera DN, Wolfswinkel L van, Lindsay TF, Beattie WS: Association between Intraoperative Hypotension and Myocardial Injury after Vascular Surgery. Anesthesiology 2016; 124:35–44
6. Monk TG, Bronsert MR, Henderson WG, Mangione MP, Sum-Ping STJ, Bentt DR, Nguyen JD, Richman JS, Meguid RA, Hammermeister KE: Association between Intraoperative Hypotension and Hypertension and 30-day Postoperative Mortality in Noncardiac Surgery. Anesthesiology 2015; 123:307–19
7. Walsh M, Devereaux PJ, Garg AX, Kurz A, Turan A, Rodseth RN, Cywinski J, Thabane L, Sessler DI: Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: toward an empirical definition of hypotension. Anesthesiology 2013; 119:507–15
8. Weiss M, Vutskits L, Hansen TG, Engelhardt T: Safe Anesthesia For Every Tot - The SAFETOTS initiative. Curr Opin Anaesthesiol 2015; 28:302–7
9. Nafiu OO, Maclean S, Blum J, Kheterpal S, Cowan A, Tremper KK: High BMI in children as a risk factor for intraoperative hypotension. Eur J Anaesthesiol 2010; 27:1065–8
10. Nafiu OO, Kheterpal S, Morris M, Reynolds PI, Malviya S, Tremper KK: Incidence and risk factors for preincision hypotension in a noncardiac pediatric surgical population. Paediatr Anaesth 2009; 19:232–9
11. Friesen RH, Wurl JL, Friesen RM: Duration of preoperative fast correlates with arterial blood pressure response to halothane in infants. Anesth Analg 2002; 95:1572–6

12. Chidambaran V, Costandi A, D'Mello A: Propofol: a review of its role in pediatric anesthesia and sedation. CNS Drugs 2015; 29:543–63
13. Naguib M, Samarkandi AH, Bakhamees HS, Magboul MA, El-Bakry AK: Histamine-release haemodynamic changes produced by rocuronium, vecuronium, mivacurium, atracurium and tubocurarine. Br J Anaesth 1995; 75:588–92
14. Wolf AR: Effects of regional analgesia on stress responses to pediatric surgery. Paediatr Anaesth 2012; 22:19–24
15. Holte K, Foss NB, Svensén C, Lund C, Madsen JL, Kehlet H: Epidural anesthesia, hypotension, and changes in intravascular volume. Anesthesiology 2004; 100:281-6
16. Rubin DB: Multiple Imputation for Nonresponse in Surveys. Edited by Rubin DB. Hoboken, NJ, USA, John Wiley & Sons, Inc., 1987
17. Donders ART, Heijden GJMG van der, Stijnen T, Moons KGM: Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006; 59:1087–91
18. Buuren S van, Groothuis-Oudshoorn K: mice : Multivariate Imputation by Chained Equations in R. J Stat Softw 2011; 45:1–67
19. Weber F, Koning L, Scoones GP: Defining hypotension in anesthetized infants by individual awake blood pressure values: a prospective observational study. Pediatr Anesth 2017
20. Dennhardt N, Beck C, Huber D, Sander B, Boehne M, Boethig D, Leffler A, Sümpelmann R, Veyckemans F: Optimized preoperative fasting times decrease ketone body concentration and stabilize mean arterial blood pressure during induction of anesthesia in children younger than 36 months: a prospective observational cohort study. Paediatr Anaesth 2016; 26:838–43
21. Bijker JB, Klei WA van, Kappen TH, Wolfswinkel L van, Moons KGM, Kalkman CJ: Incidence of intraoperative hypotension as a function of the chosen definition: literature definitions applied to a retrospective cohort using automated data collection. Anesthesiology 2007; 107:213–20
22. Vernooij LM, Klei WA van, Machina M, Pasma W, Beattie WS, Peelen LM: Different methods of modelling intraoperative hypotension and their association with postoperative complications in patients undergoing non-cardiac surgery. Br J Anaesth 2018; 120:1080–9
23. Hoorweg A-LJ, Pasma W, Wolfswinkel L van, Graaff JC de: Incidence of Artifacts and Deviating Values in Research Data Obtained from an Anesthesia Information Management System in Children. Anesthesiology 2018; 128:293–304

Supplemental Material

S1: Relevant period for height measurements

Age	Relevant period
0 to 1 months	7 days
1 to 3 months	14 days
6 to 12 months	30 days
1 to 4 years	60 days
>4 years	90 days

Height measurements were collected, within a relevant period before surgery. When there was no height measurement available within this period, height was considered missing.


II

Artifacts in physiologic anesthesia data

4

Incidence of Artifacts and Deviating Values in Research Data Obtained from an Anesthesia Information Management System in Children

Anne-lee J. Hoorweg
Wietze Pasma
Leo van Wolfswinkel
Jurgen C. de Graaff

Published in Anesthesiology 2018 Feb;128(2):293-304

Abstract

Background

Vital parameter data collected in anesthesia information management systems are often used for clinical research. The validity of this type of research is dependent on the number of artifacts.

Methods

In this prospective observational cohort study, the incidence of artifacts in anesthesia information management system data was investigated in children undergoing anesthesia for noncardiac procedures. Secondary outcomes included the incidence of artifacts among deviating and nondeviating values, among the anesthesia phases, and among different anesthetic techniques.

Results

We included 136 anesthetics representing 10,236 min of anesthesia time. The incidence of artifacts was 0.5% for heart rate (95% CI: 0.4 to 0.7%), 1.3% for oxygen saturation (1.1 to 1.5%), 7.5% for end-tidal carbon dioxide (6.9 to 8.0%), 5.0% for noninvasive blood pressure (4.0 to 6.0%), and 7.3% for invasive blood pressure (5.9 to 8.8%). The incidence of artifacts among deviating values was 3.1% for heart rate (2.1 to 4.4%), 10.8% for oxygen saturation (7.6 to 14.8%), 14.1% for end-tidal carbon dioxide (13.0 to 15.2%), 14.4% for noninvasive blood pressure (10.3 to 19.4%), and 38.4% for invasive blood pressure (30.3 to 47.1%).

Conclusions

Not all values in anesthesia information management systems are valid. The incidence of artifacts stored in the present pediatric anesthesia practice was low for heart rate and oxygen saturation, whereas noninvasive and invasive blood pressure and end-tidal carbon dioxide had higher artifact incidences. Deviating values are more often artifacts than values in a normal range, and artifacts are associated with the phase of anesthesia and anesthetic technique. Development of (automatic) data validation systems or solutions to deal with artifacts in data is warranted.

Introduction

Anesthesia information management systems (AIMSs) are increasingly being used for anesthesiologic recordkeeping.[1] Electronic patient records are considered to be better than handwritten anesthesia records because they require less time and are more complete, accurate, and reliable.[2–7] AIMSs provide improved recordkeeping of anesthetic procedures, accurate guiding of patient management, and enhanced patient safety, because the anesthesiologist can focus on intraoperative events instead of manual charting.[1,7,8] Additionally, data from AIMSs are a valuable resource for database research[9–11] and medicolegal litigations.[12–14]

Although the quality of data capturing and registering in AIMSs is considered highly accurate, not all stored vital parameter values are based on valid measurements. The course of anesthesia often clarifies the accuracy of a measurement in the clinical situation. However, during data acquisition from the monitoring system, the AIMS cannot verify whether a particular value is a true value or an artifact. As a consequence, retrospectively it is difficult or impossible to differentiate between values representing the actual patient's vital state and artifacts. Hence, artifacts influence the validity of research based on AIMSs. Furthermore, artifacts may lead to bias, if certain artifacts are associated with specific patient characteristics (e.g., age) or the phase of anesthesia.

Previous research demonstrated a low incidence of artifacts in an AIMS database of an adult population; e.g., 0.3% of all oxygen saturation measured by pulse oximetry (SpO2) values and 24.2% of deviating SpO2 values were caused by artifacts.[15] In children, the incidence of artifacts has thus far only been investigated for SpO2 and was higher than in adults; 46% of episodes with an SpO2 of at most 90% occurring in children under anesthesia were artifacts.[11]

We hypothesized that the difference in physiology and anesthetic technique between children and adults would result in a different incidence of artifacts and deviating values. Therefore, we assessed the validity of AIMS data in a pediatric population undergoing a procedure under general anesthesia and defined factors associated with artifacts.

Materials and Methods

Setting and Study Population

The study was approved by the local Medical Research Ethics Committee of the University Medical Center Utrecht (Utrecht, The Netherlands), which waived the need for informed consent, because subjects were not exposed to a research intervention. According to the requirements of Dutch law, anonymity and confidentiality of routinely collected clinical data were assured.

In this prospective observational cohort study, we included pediatric patients who underwent general anesthesia for noncardiac pediatric surgical or diagnostic procedures in a tertiary pediatric university hospital (Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands) between May and August 2015. One operating room per day was randomly assigned by drawing paper lots with numbers corresponding with the available operating rooms in sealed envelopes. There were five operating rooms available for randomization. Operating rooms were excluded when there were only cardiothoracic and/or angiographic procedures performed, because of possible preexistent abnormal vital parameter values and/or expected long periods of unreliable vital parameter measurements or when they were located outside the operating complex (e.g., magnetic resonance imaging suite). All anesthetics were observed by the investigator (A.-l.J.H.), a resident anesthesiologist, who was trained for this research by a specialized pediatric anesthesiologist (J.C.d.G.) to identify artifacts. Because a child could be included more than once, the results are reported in relation to the number of anesthetics. An anesthetic was defined as the registration of an anesthetic procedure from induction to the end of anesthesia when the patient left the operating room. All anesthetics were performed by specialized pediatric anesthesiologists.

Patients were monitored during anesthesia with heart rate (HR), SpO2, noninvasive blood pressure, and (not in every patient) invasive blood pressure by an IntelliVue monitoring system (type MP70, X2 multimeasurement module; Philips, Germany) with a built-in filter for artifacts. The HR displayed on the monitor was derived, according to availability, from the electrocardiogram and/or the plethysmograph; when both of these HR values were available, they were displayed next to each other. The HR value was updated after every new measured QRS complex by calculating the mean HR over the last twelve R–R intervals when the HR was at

least 50 beats/min or over the last four R–R intervals when the HR was less than 50 beats/min or by averaging the detected arterial pulsations over the last 8 s, respectively. SpO2 was measured by pulse oximetry (Philips FAST) with a taped sensor (OxiMax-P, single patient use adhesive sensor; Covidien-Nellcor, USA).[16] SpO2 was displayed on the monitor as the median value over the last 5 s with an update period of 2 s. The mean noninvasive blood pressure was measured by oscillometry and was displayed on the monitor each time it was measured (usually the interval was set at every 5 min).[17] The mean invasive blood pressure was derived from the invasive blood pressure curve, which was displayed beat to beat, and was calculated over the last eight beat pressure values.

End-tidal carbon dioxide (ETCO2) was measured by a Cicero anesthesia ventilator (Dräger, Germany) by sidestream sampling of carbon dioxide (200 ml/min) in exhaled gas. Carbon dioxide was detected by infrared spectroscopy, and the ETCO2 of each breath was displayed on the Dräger ventilator and with a delay of 9 s displayed on the Philips monitor.[18]

Values from the Philips monitoring system were stored in a locally developed AIMS (AnStat, version 2.0.4, 2015; Carepoint, The Netherlands) that automatically samples data from the monitoring system every 5 s.[11,15,19] This AIMS has a low-pass filter that records data in the database every min after filtering has been applied. This filtering implies that the median value per min was calculated and stored for HR, SpO2, and mean invasive blood pressure and that the highest value per min was stored for ETCO2. The mean noninvasive blood pressure was recorded every time it was measured without filtering. The HR can be sampled from the electrocardiogram, plethysmograph, and invasive blood pressure in the Philips IntelliVue monitoring system. The HR recorded in the AIMS was primarily derived from the electrocardiogram, but if this value was not available from the monitor, the HR was derived from the plethysmograph. The origin of the HR value was not stored in the AIMS, only the value. Storing the median (HR, SpO2, and mean invasive blood pressure) or highest (ETCO2) value/min was considered an effective method to filter for the majority of artifacts, because, compared to the mean value, the median respectively highest value of the 12 values captured per min is not influenced by short-lasting artifacts.[11,15]

Data Collection

All data for HR, SpO2, ETCO2, and mean noninvasive and invasive blood pressure were collected automatically by the AIMS and visualized on a computer screen in the operating room. The data were collected according to standard clinical practice from the time the monitoring system was connected to the patient before induction until the monitor was disconnected when the patient left the operating room. The investigator (A.-l.J.H.) was present in the operating room during all included procedures and compared the monitor with the values that were stored in the AIMS. Each value in the AIMS was inspected and assessed as being an artifact or a valid value. The phase of anesthesia during artifact occurrence and the causes of artifacts were also documented in the operating room. After data collection, the AIMS database was queried to add additional information to the collected data, including the anesthetic technique and whether a value was deviating from a predefined reference range.

The data collection only took into account data that were stored and displayed in the AIMS. Therefore, data filtered out by the Philips monitoring system or by the AIMS were not taken into consideration. An artifact was defined as any value that was judged invalid and/or not reflecting the patient's current physiologic state, based on the investigator's (A.-l.J.H.) consultation with the attending anesthesiologist regarding measurements, physiologic state, and observations in the operating room.

In case of discrepancy between the investigator and the attending anesthesiologist, a second investigator (J.C.d.G., pediatric anesthesiologist and primary investigator) made the final decision. The primary rater (A.-l.J.H.) was trained in the specific observation skills by the second investigator (J.C.d.G.) during a period of 2 weeks before the start of the study.

A deviating value was defined as a value outside a predefined reference range. For SpO2[11,20,21] and ETCO2,[22–24] fixed reference ranges were used (table 1). ETCO2 reference values were calculated from reference values for arterial carbon dioxide, ranging from 35 to 45 mmHg. ETCO2 is normally 2 to 5 mmHg lower than arterial carbon dioxide because of mixing of carbon dioxide containing alveolar gas with expired gas empty of carbon dioxide from the anatomical dead space. We calculated ETCO2 reference values based on this difference.[23,24] For HR,

reference values were based on age.[25] For mean blood pressure, reference values were based on age and sex.[26] We focused on mean noninvasive and invasive blood pressure, because mean blood pressure is considered to be most important under anesthesia in daily care in our hospital.

Table 1: Definition of Deviating Values

Parameter	Deviating Value	
Heart rate	< 5th percentile*	> 95th percentile*
SpO2	≤ 90% [11,20,21]	NA
ETCO2	< 30 mmHg [22–24]	> 43 mmHg [22–24]
Mean NIBP	< –2SD†	> +2SD†
Mean IBP	< –2SD†	> +2SD†

*In the absence of available reference values derived in children under anesthesia, age-dependent cutoff values for heart rate were based on data acquired in two tertiary-care children's hospitals in children on general medical and surgical wards.[25] †Age- and sex-dependent cutoff values for mean NIBP and IBP were based on research data acquired in a multicenter study in children under anesthesia.[26] ETCO2 = end-tidal carbon dioxide; IBP = invasive blood pressure; NA = not applicable; NIBP = noninvasive blood pressure; SpO2 = oxygen saturation.

Outcomes

The primary outcome was the incidence of artifacts for each included vital parameter. Secondary outcomes comprised the incidence of deviating values, and artifacts as a proportion of deviating and nondeviating values. Additionally, we made group comparisons, determining the artifact incidence among the three anesthesia phases and among different age categories. In addition, the causes of artifacts were recorded. The incidences of ETCO2 artifacts and deviating values were related to the anesthetic technique in a post hoc analysis.

Sample Size and Statistical Analysis

To perform a sample size calculation, we predefined an artifact incidence of less than 5% to yield valid AIMS data.[15,27] We assumed a maximum 4% incidence of artifacts, and we aimed for a 95% CI from 3 to 5% (lower to upper limit), because less than 5% has been considered a valid artifact percentage for AIMS data.[15,27] Moreover, for this calculation, we assumed that artifacts within patients were independent. Because mean noninvasive blood pressure is the measurement performed least frequently, the sample size calculation was based on the hypothesis

that 4% of the recorded mean noninvasive blood pressure values would be an artifact. The additional assumptions that the mean noninvasive blood pressure would be measured at least every 5 min and a 95% CI would not exceed 5% led to a required minimum of 1,825 mean noninvasive blood pressure measurements or 9,125 min of anesthesia time.

To assess the association between the artifact incidence and the factors studied, a mixed effects model was constructed per determinant (deviating status of measurement, anesthesia phase, age category, or anesthetic technique), including a random intercept per case. This model was compared with an empty model without the determinant, including a random intercept per case, with a likelihood ratio test. Both binomial mixed effects models use a logit link function and were fitted using maximum likelihood estimation. An association between the determinant and the artifact incidence was considered statistically significant if the likelihood ratio test showed a p-value of at most 0.05. Statistical analysis was performed with SPSS version 22.0 (IBM, USA) and R version 3.3.2 (R core team 2016).

Results

In this study, we included 136 anesthetics in 132 patients during a cumulative time of anesthesia of 10,236 min (table 2). HR, SpO2, and ETCO2 were measured in all anesthetics, whereas mean noninvasive blood pressure was not measured in eight anesthetics (mostly short ear-nose-throat procedures), and mean invasive blood pressure was measured in five anesthetics (table 3). The percentage of artifacts was lowest for HR (0.5%; 95% CI: 0.4 to 0.7%) and SpO2 (1.3%; 1.1 to 1.5%; table 3; figs. 1 and 2). For mean noninvasive blood pressure, the artifact incidence was 5.0% (4.0 to 6.0%). Mean invasive blood pressure (7.3%; 5.9 to 8.8%) and ETCO2 (7.5%; 6.9 to 8.0%) showed higher incidences. ETCO2 values contained the most deviating values (40.7%; 95% CI: 39.7 to 41.7%), whereas only 3.1% (2.8 to 3.5%) of SpO2 values deviated from the predefined reference range. Deviating values were more often artifacts than values in a normal range ($P < 0.001$ for every vital parameter), varying between 3.1% (95% CI: 2.1 to 4.4%; HR) and 38.4% (30.3 to 47.1%; mean invasive blood pressure) artifacts among deviating values and between 0.2% (0.1 to 0.3%; HR) and 3.5% (2.7 to 4.5%; mean noninvasive blood pressure) artifacts among nondeviating values (table 3; figs. 1 and 2). Almost all anesthetics showed ETCO2 artifacts (94.1%), unlike HR artifacts, which occurred in only 15 (11.0%) of the

anesthetics. Only 23.5% of anesthetics demonstrated deviations in SpO2, contrary to more than 50% of anesthetics for all other vital parameters.

Table 2: Baseline Characteristics

	All anesthetics*		**Median anesthesia time in minutes (IQR)**		**Total anesthesia time in minutes**
	***n* = 136**		**58 (35-94)**		**10,236**
Age					
0-29 days	4	(2.9)	186	(45-380)	849
29 days – 1 year	25	(18.4)	66	(48-131)	2,152
1-4 years	27	(19.9)	53	(21-71)	1,461
4-12 years	52	(38.2)	53	(28-83)	3,286
12-18 years	28	(20.6)	63	(47-113)	2,488
Gender					
Male	78	(57.4)	61	(43-98)	6,137
Female	58	(42.6)	55	(28-89)	4,099
Procedure type					
General surgery	31	(22.8)	67	(44-96)	2,851
General pediatric procedures†	25	(18.4)	40	(18-59)	1,045
Urology	24	(17.6)	61	(46-116)	1,984
Ear-nose-throat surgery	19	(14.0)	33	(12-65)	970
Plastic and reconstructive surgery	14	(10.3)	105	(55-137)	1,322
Orthopedic surgery	9	(6.6)	85	(57-142)	1,019
Dental surgery	5	(3.7)	63	(58-70)	373
Ophthalmology	5	(3.7)	39	(24-48)	184
Neurosurgery	4	(2.9)	106	(78-167)	488
Anesthesia induction type					
Inhalation (sevoflurane)	86	(63.2)	59	(38-98)	6,312
Intravenous (propofol)	50	(36.8)	57	(29-85)	3,924

Table 2 *(continued)*: Baseline Characteristics

	All anesthetics*		**Median anesthesia time in minutes (IQR)**		**Total anesthesia time in minutes**
	***n* = 136**		**58 (35-94)**		**10,236**
Anesthesia maintenance type					
Inhalation (sevoflurane) with tube	61	(44.9)	83	(59-131)	6,122
Inhalation (sevoflurane) with LMA	52	(38.2)	48	(32-62)	2,791
Inhalation (sevoflurane) with mask	16	(11.8)	16	(12-25)	329
Intravenous (propofol)	7	(5.1)	85	(60-222)	994
Airway device type					
Tube					
Cuffed	58	(42.6)	83	(58-132)	5,859
Uncuffed	4	(2.9)	113	(79-214)	724
Combination cuffed/ uncuffed	1	(0.7)	137		137
LMA	59	(43.4)	50	(33-62)	3,236
Mask	14	(10.3)	16	(11-26)	280
Patient position					
Supine	101	(74.3)	58	(38-89)	6,965
Lithotomy	15	(11.0)	56	(40-64)	955
Lateral	15	(11.0)	52	(18-132)	1,553
Combination	3	(2.2)	84	(56-120)	269
Prone	2	(1.5)	247	(205-289)	494

*All values in the second column are reported as numbers (%). An anesthetic was defined as the registration of an anesthetic procedure from induction to the end of anesthesia when the patient left the operating room. †This includes procedures such as a bone marrow puncture, lumbar puncture, or duodenoscopy. IQR = interquartile range; LMA = laryngeal mask airway.

Table 3: Results of Analysis Performed on All Values during Anesthesia

All anesthetics (n = 136)* Parameter	Observed Anesthesia time (minutes)	Number of values	Number of values being artifacts (%; 95% CI)		Number of anesthetics with artifacts (%)*		Number of values being deviating values (%; 95% CI)		Number of anesthetics with deviating values (%)*		% of deviating values being artifacts (95% CI)		% of non-deviating values being artifacts (95% CI)	
Heart rate	10,236	10,076	50	(0.5; 0.4-0.7)	15	(11.0)	1,003	(10.0; 9.4-10.6)	99	(72.8)	3.1	(2.1-4.4)	0.2	(0.1-0.3)†
SpO2	10,236	10,020	130	(1.3; 1.1-1.5)	53	(39.0)	315	(3.1; 2.8-3.5)	32	(23.5)	10.8	(7.6-14.8)	1.0	(0.8-1.2)†
ETCO2	10,236	9,720	725	(7.5; 6.9-8.0)	128	(94.1)	3,955	(40.7; 39.7-41.7)	135	(99.3)	14.1	(13.0-15.2)	2.9	(2.5-3.4)†
Mean NIBP	10,132	1,878	93	(5.0; 4.0-6.0)	33	(25.8)	250	(13.3; 11.8-14.9)	67	(52.3)	14.4	(10.3-19.4)	3.5	(2.7-4.5)†
Mean IBP	1,442	1,255	91	(7.3; 5.9-8.8)	5	(100)	138	(11.0; 9.3-12.9)	4	(80.0)	38.4	(30.3-47.1)	3.4	(2.4-4.6)†

Observed anesthesia time represents the cumulative number of min the patients were under anesthesia. Heart rate, SpO2, and ETCO2 have the same anesthesia time; mean NIBP and IBP times are lower, because mean NIBP was not measured in eight anesthetics and mean IBP was only measured in five anesthetics. The number of values for heart rate, SpO2, ETCO2, and IBP differ slightly from anesthesia time, although these parameters are calculated and registered every min; the difference is caused by some missing parameters in the anesthesia information management system. The number of mean NIBP values is lower than the anesthesia time, because the NIBP measurement interval was usually set at every 5min.

*An anesthetic was defined as the registration of an anesthetic procedure from induction to the end of anesthesia when the patient left the operating room. †Statistical significance ($P < 0.001$) between the percentage of deviating values being artifacts and the percentage of nondeviating values being artifacts.

ETCO2 = end-tidal carbon dioxide; IBP = invasive blood pressure; NIBP = noninvasive blood pressure; SpO2 = oxygen saturation.

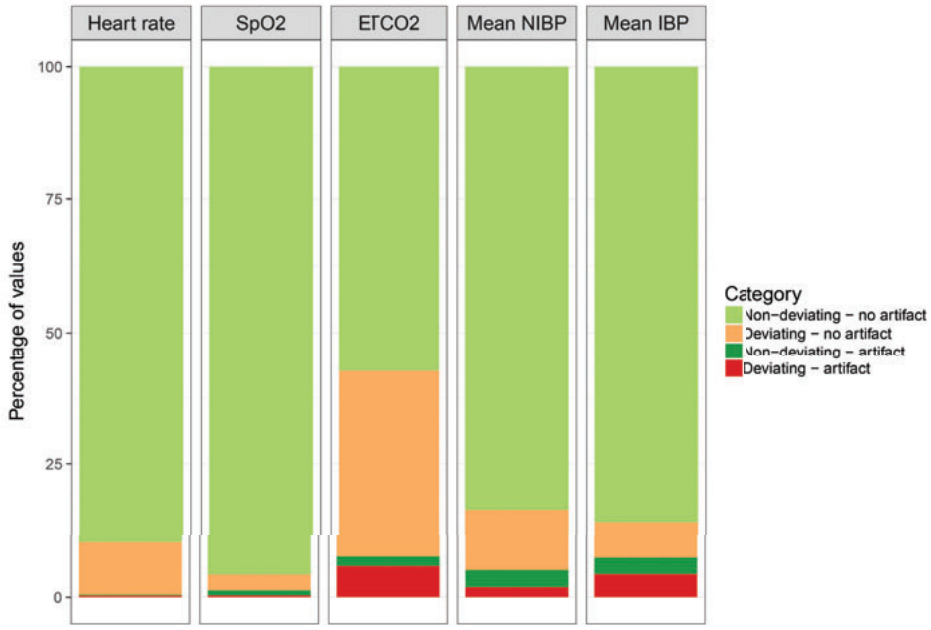


Figure 1: Overview of artifacts and deviating values. This overview shows that the percentages of artifacts and deviating values for heart rate and oxygen saturation measured by pulse oximetry (SpO2) were relatively low. End-tidal carbon dioxide (ETCO2) values most often deviated outside the predefined reference range, and these values also contained relatively many artifacts. Mean noninvasive blood pressure (NIBP) and invasive blood pressure (IBP) demonstrated the most nondeviating values being artifacts.

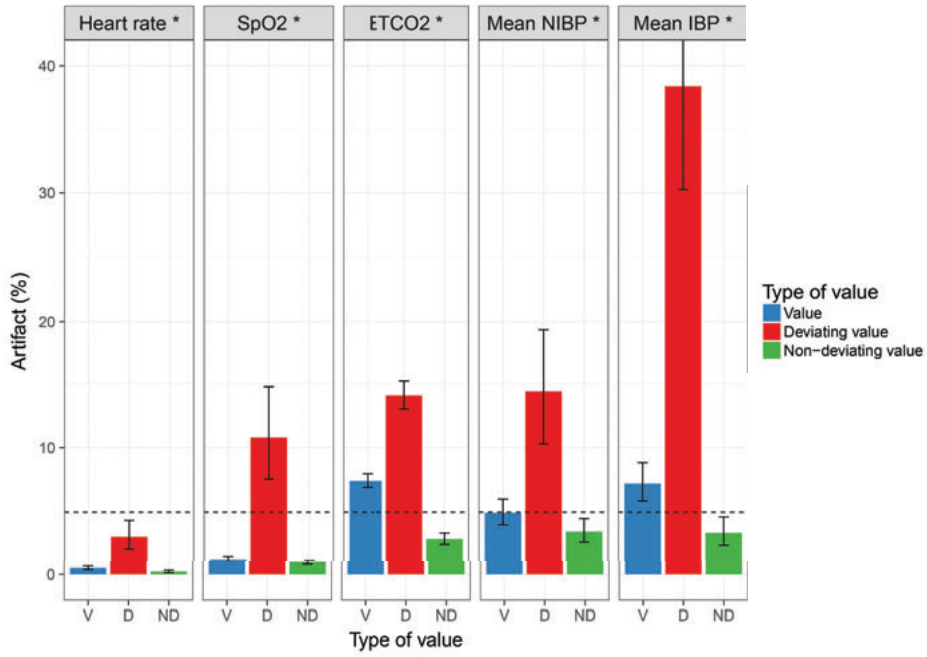


Figure 2: Artifacts among all values and among deviating and nondeviating values. The percentage of artifacts was lowest for heart rate (0.5%) and oxygen saturation measured by pulse oximetry (SpO2; 1.3%). For mean noninvasive blood pressure (NIBP), the artifact incidence was 5.0%. Mean invasive blood pressure (IBP; 7.3%) and end-tidal carbon dioxide (ETCO2; 7.5%) showed higher incidences of artifacts. Deviating values (D) were more often artifacts than nondeviating values (ND), which was statistically significant for all vital parameters (marked by asterisks). The dashed line indicates the predefined threshold for valid data (5.0% artifacts). The whiskers show 95% CI. V = value.

For HR, the artifact incidence was higher during the induction phase (1.0%) compared to anesthesia maintenance and emergence ($P < 0.001$; table 4; figure 3). For SpO2, ETCO2, and mean invasive blood pressure, the artifact incidence was higher during the induction and emergence phases compared to anesthesia maintenance, with the highest incidence for SpO2 (4.6%) and ETCO2 (23.6%) during induction (both $P < 0.001$) and for mean invasive blood pressure during emergence (37.6%; $P < 0.001$). For mean noninvasive blood pressure, the artifact incidence was not significantly associated with the anesthesia phases.

Table 4: Artifacts per Phase of Anesthesia

Parameter (number of artifacts)	Phase of anesthesia	Number of values	Number of values being artifacts	Artifacts as % of artifacts per parameter*	Artifact incidence (%; 95% CI) †		P value association
Heart rate (n = 50)	Induction	1,833	18	36.0	1.0	(0.6-1.6)	< 0.001
	Maintenance	7,073	27	54.0	0.4	(0.3-0.6)	
	Emergence	1,170	5	10.0	0.4	(0.1-1.0)	
SpO2 (n = 130)	Induction	1,814	83	63.8	4.6	(3.7-5.6)	< 0.001
	Maintenance	7,039	15	11.5	0.2	(0.1-0.4)	
	Emergence	1,167	32	24.6	2.7	(1.9-3.9)	
ETCO2 (n = 725)	Induction	1,536	362	49.9	23.6	(21.5-25.8)	< 0.001
	Maintenance	7,063	136	18.8	1.9	(1.6-2.3)	
	Emergence	1,121	227	31.3	20.2	(17.9-22.7)	
Mean NIBP (n = 93)	Induction	332	19	20.4	5.7	(3.5-8.8)	0.337
	Maintenance	1,450	70	75.3	4.8	(3.8-6.1)	
	Emergence	96	4	4.3	4.2	(1.2-10.3)	
Mean IBP (n = 91)	Induction	203	27	29.7	13.3	(9.0-18.8)	< 0.001
	Maintenance	919	14	15.4	1.5	(0.8-2.5)	
	Emergence	133	50	54.9	37.6	(29.4-46.4)	

*Percentages represent the distribution of artifacts among the three anesthesia phases per vital parameter. †Percentages represent the artifact incidence during a specific anesthesia phase per vital parameter. ETCO2 = end-tidal carbon dioxide; IBP = invasive blood pressure; NIBP = noninvasive blood pressure; SpO2 = oxygen saturation.

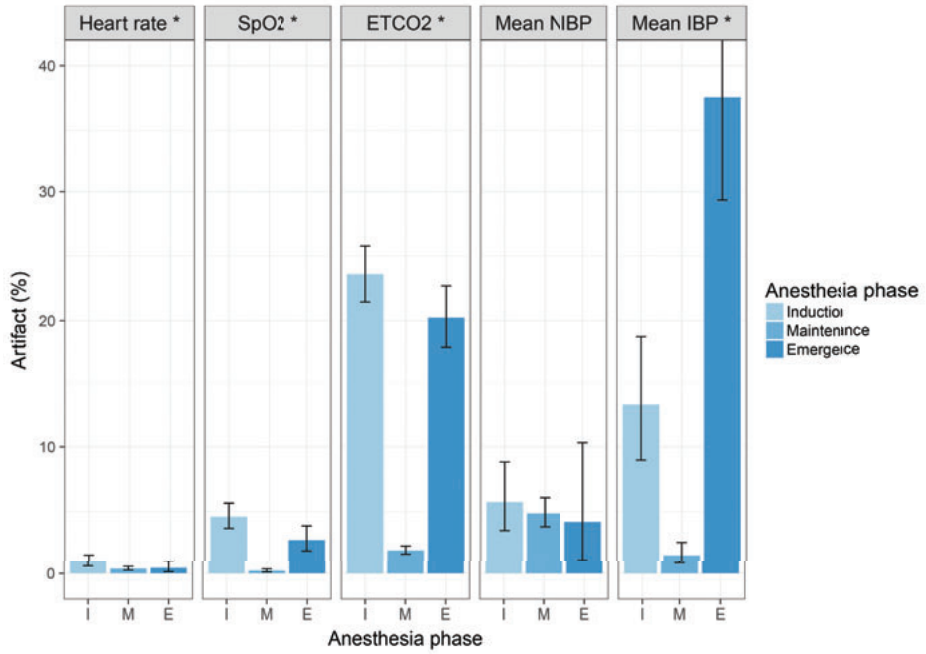


Figure 3: Artifacts among anesthesia phases. For heart rate, the incidence of artifacts was higher during the induction (I) phase (1.0%) compared to anesthesia maintenance (M) and emergence (E). For oxygen saturation measured by pulse oximetry (SpO2), end-tidal carbon dioxide (ETCO2), and mean invasive blood pressure (IBP), the artifact incidence was higher during the induction and emergence phases compared to anesthesia maintenance, with the highest incidence for SpO2 (4.6%) and ETCO2 (23.6%) during induction and for mean IBP during emergence (37.6%). For mean noninvasive blood pressure (NIBP), the artifact incidence was not significantly different between the three anesthesia phases. The whiskers show 95% CI. The asterisks indicate statistically significant differences between the anesthesia phases.

ETCO2 artifacts showed a higher incidence in children up to 4 yr of age (10.3 to 11.3%) compared to older children (4.0 to 5.1%; P = 0.001; table 5; figure 4). Artifact incidences for the other parameters did not differ significantly between age groups. The artifact incidence was low in all age groups for HR (range: 0.1 to 1.3%; P = 0.170) and SpO2 (1.0 to 3.1%; P = 0.130). The most common causes of artifacts were electrocautery for HR (52.0%), patient movement for SpO2 (40.0%), mask ventilation during induction for ETCO2 (40.4%), an oversized pressure cuff for mean noninvasive blood pressure (37.6%), and relocation of the pressure sensor for mean invasive blood pressure (24.2%; table 6).

Table 5: Artifacts per Age Group

Parameter	Age group	Observed anesthesia time (minutes)	Number of values	Number of values being artifacts (%; 95% CI)	
Heart rate (n=50)	0-29 days	849	828	11	(1.3; 0.7-2.4)
	29 days – 1 yr	2,152	2,150	4	(0.2; 0.1-0.5)
	1-4 yr	1,461	1,412	6	(0.4; 0.2-0.9)
	4-12 yr	3,286	3,199	3	(0.1; 0.0-0.3)
	12-18 yr	2,488	2,487	26	(1.0; 0.7-1.5)
SpO2 (n = 130)	0-29 days	849	828	26	(3.1; 2.1-4.6)
	29 days – 1 yr	2,152	2,145	24	(1.1; 0.7-1.7)
	1-4 yr	1,461	1,411	14	(1.0; 0.5-1.7)
	4-12 yr	3,286	3,186	33	(1.0; 0.7-1.5)
	12-18 yr	2,488	2,450	33	(1.3; 0.9-1.9)
ETCO2 (n = 725)*	0-29 days	849	766	79	(10.3; 8.3-12.7)
	29 days – 1 yr	2,152	2,103	238	(11.3; 10.0-12.7)
	1-4 yr	1,461	1,382	156	(11.3; 9.7-13.1)
	4-12 yr	3,286	3,076	157	(5.1; 4.4-5.9)
	12-18 yr	2,488	2,393	95	(4.0; 3.2-4.8)
Mean NIBP (n = 93)	0-29 days	849	146	3	(2.1; 0.4-5.9)
	29 days – 1 yr	2,152	407	7	(1.7; 0.7-3.5)
	1-4 yr	1,415	275	8	(2.9; 1.3-5.7)
	4-12 yr	3,254	657	56	(8.5; 6.5-10.9)
	12-18 yr	2,462	393	19	(4.8; 2.9-7.4)
Mean IBP (n = 91)	0-29 days	759	751	62	(8.3; 6.4-10.5)
	29 days – 1 yr†	-	-		-
	1-4 yr†	-	-		-
	4-12 yr	149	18	3	(16.7; 3.6-41.4)
	12-18 yr	534	486	26	(5.3; 3.5-7.7)

Observed anesthesia time represents the cumulative number of min the patients were under anesthesia. Heart rate, SpO2, and ETCO2 have the same anesthesia time; mean NIBP and IBP times are lower, because mean NIBP was not measured in eight anesthetics, and mean IBP was measured in only five anesthetics. The number of values for heart rate, SpO2, ETCO2, and IBP differ slightly from anesthesia time, although these parameters are calculated and registered every min; the difference is caused by some missing parameters in the anesthesia information management system. The number of mean NIBP values is lower than the anesthesia time, because the NIBP measurement interval was usually set at every 5min.

*Indicates statistical significance (P = 0.001) between age groups for ETCO2. †No children between 29 days and 4 yr were monitored with IBP.

ETCO2 = end-tidal carbon dioxide; IBP = invasive blood pressure; NIBP = noninvasive blood pressure; SpO2 = oxygen saturation.

Table 6: Causes of Artifacts per Age Group

Parameter (number of artifacts)	Cause	All anesthetics (*n* = 136)*		0-29 days (*n* = 4)		29 days – 1 year (*n* = 25)		1-4 years (*n* = 27)		4-12 years (*n* = 52)		12-18 years (*n* = 28)	
Heart rate (*n* = 50)	Electrocautery	26	(52.0)	0		0		0		0		26	(100)
	Patient not connected to ECG, pulse oximeter or arterial line	17	(34.0)	10	(90.9)	3	(75.0)	2	(33.3)	2	(66.7)	0	
	Patient movement	2	(4.0)	0		1	(25.0)	1	(16.7)	0		0	
	Dislodgement of ECG electrode(s)	1	(2.0)	0		0		0		1	(33.3)	0	
	Unknown	4	(8.0)	1	(9.1)	0		3	(50.0)	0		0	
SpO2 (*n* =130)	Patient movement	52	(40.0)	7	(26.9)	13	(54.2)	7	(50.0)	15	(45.5)	10	(30.3)
	Patient not connected to pulse oximeter	33	(25.4)	8	(30.8)	3	(12.5)	4	(28.6)	7	(21.2)	11	(33.3)
	Low peripheral perfusion†	23	(17.7)	1	(3.8)	7	(29.2)	1	(7.1)	9	(27.3)	5	(15.2)
	Peripheral vasoconstriction‡	17	(13.1)	10	(38.5)	0		0		2	(6.1)	5	(15.2)
	Dislocated pulse oximeter	1	(0.8)	0		1	(4.2)	0		0		0	
	Unknown	4	(3.1)	0		0		2	(14.3)	0		2	(6.1)
ETCO2 (*n* = 725)	Mask ventilation during induction	293	(40.4)	7	(8.9)	86	(36.1)	75	(48.1)	80	(51.0)	45	(47.4)
	Patient not connected to anesthesia ventilator§	198	(27.3)	19	(24.1)	70	(29.4)	38	(24.4)	48	(30.6)	23	(24.2)
	Mask ventilation during emergence	83	(11.4)	12	(15.2)	24	(10.1)	16	(10.3)	14	(8.9)	17	(17.9)
	Mask ventilation during procedure (no tube or LMA)	46	(6.3)	7	(8.9)	18	(7.6)	9	(5.8)	9	(5.7)	3	(3.2)
	Gas leak alongside tube or LMA	43	(5.9)	0		31	(13.0)	7	(4.5)	2	(1.3)	3	(3.2)
	Rapid respiration (small tidals) ||	36	(5.0)	34	(43.0)	1	(0.4)	1	(0.6)	0		0	
	Airway obstruction#	12	(1.7)	0		2	(0.8)	8	(5.1)	2	(1.3)	0	
	Manual ventilation	7	(1.0)	0		5	(2.1)	0		0		2	(2.1)
	Airway device obstruction#	3	(0.4)	0		1	(0.4)	1	(0.6)	0		1	(1.1)
	Apnea#	1	(0.1)	0		0		0		0		1	(1.1)
	Unknown	3	(0.4)	0		0		1	(0.6)	2	(1.3)	0	

Table 6 *(continued)*: Causes of Artifacts per Age Group

Parameter (number of artifacts)	Cause	All anesthetics (*n* = 136)*		0-29 days (*n* = 4)		29 days – 1 year (*n* = 25)		1-4 years (*n* = 27)		4-12 years (*n* = 52)		12-18 years (*n* = 28)	
Mean NIBP (*n* = 93)	Oversized pressure cuff	35	(37.6)	0		0		0		35	(62.5)	0	
	Patient movement	22	(23.7)	0		4	(57.1)	6	(75.0)	8	(14.3)	4	(21.1)
	Abnormal positioning of arm (cuff not at heart level, e.g. above body)	14	(15.1)	0		0		0		0		14	(73.7)
	Leaning against pressure cuff by patient	11	(11.8)	0		1	(14.3)	0		10	(17.9)	0	
	Leaning against pressure cuff by other	10	(10.8)	3	(100)	2	(28.6)	2	(25.0)	2	(3.6)	1	(5.3)
	Tourniquet just distal to pressure cuff	1	(1.1)	0		0		0		1	(1.8)	0	
Mean IBP (*n* = 91)**	Relocation of pressure sensor	22	(24.2)	2	(3.2)	-		-		0		20	(76.9)
	Pressure sensor already connected, but patient not connected to sensor	21	(23.1)	21	(33.9)	-		-		0		0	
	Patient movement	17	(18.7)	17	(27.4)	-		-		0		0	
	Closed arterial line for blood sampling	16	(17.6)	15	(24.2)	-		-		0		1	(3.8)
	Dampened curve	9	(9.9)	2	(3.2)	-		-		2	(66.7)	5	(19.2)
	Manual compression of artery proximal to arterial line	5	(5.5)	5	(8.1)	-		-		0		0	
	Pressure sensor already connected, but patient without arterial line	1	(1.1)	0		-		-		1	(33.3)	0	

All values are numbers (%) of artifacts per vital parameter.

*An anesthetic was defined as the registration of an anesthetic procedure from induction to the end of anesthesia when the patient left the operating room.

†Caused by blood pressure measurement, a tourniquet for insertion of a venous cannula or arterial cannulation on the same arm or leg as the pulse oximeter or external pressure on the same arm or leg as the pulse oximeter. ‡With pulsatile flow indicator of less than 0.3 (0.3 is indicated by the manufacturer as the limit for a reliable SpO2 measurement), in the absence of causes mentioned under †, e.g., caused by stress or hypothermia. ||With rapid respiration, the system may not have sufficient time to reach 100% of the required response, leading to lower than actual ETCO2, because of interruption of the exhalation by the next inspiration. In addition, rapid shallow breathing increases dead space ventilation, leading to lower than actual ETCO2 values. §During intubation or LMA placement, during laryngotracheobronchoscopy, or at the beginning or end of a procedure. #Because of airway (device) obstruction or apnea, alveolar CO2 cannot reach the CO2 sensor, and consequently the measured ETCO2 will be low, whereas the actual alveolar CO2 is expected to be high at that moment. **No children between 29 days to 4 yr were monitored with IBP. ECG = electrocardiography; ETCO2 = end-tidal carbon dioxide; IBP = invasive blood pressure; LMA = laryngeal mask airway; NIBP = noninvasive blood pressure; SpO2 = oxygen saturation.[g] No children between 29 days – 4 years were monitored with IBP.

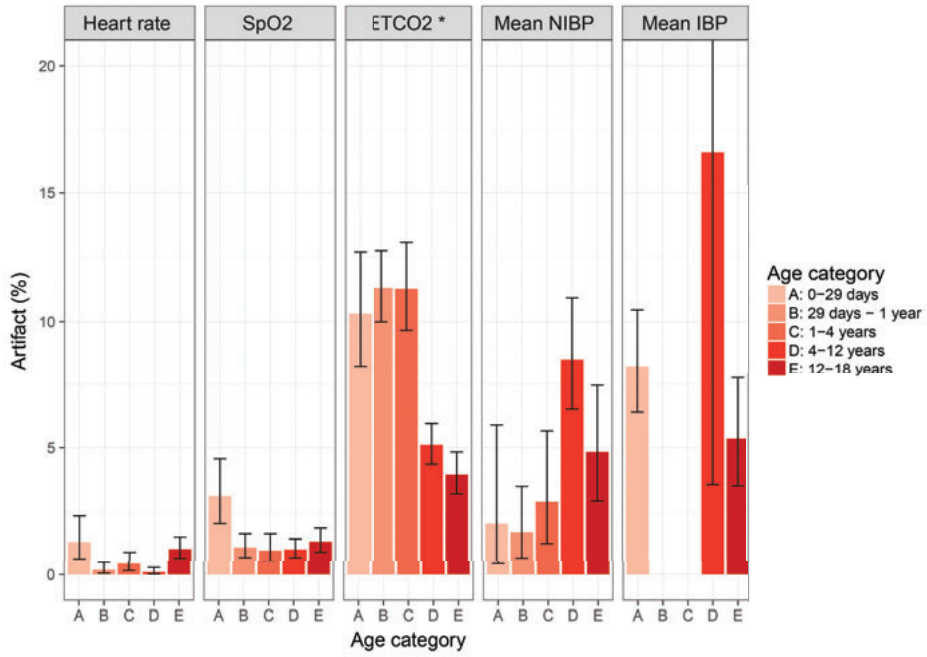


Figure 4: Artifacts among age categories. Only for end-tidal carbon dioxide (ETCO2), there were statistically significant differences between the age categories (marked by *), with a higher incidence of ETCO2 artifacts in children up to 4 yr of age (10.3 to 11.3%) compared to older children (4.0 to 5.1%). Artifact incidences for the other parameters did not differ significantly between age groups. The artifact incidence was low in all age groups for heart rate and oxygen saturation measured by pulse oximetry (SpO2). No children between 29 days and 4 yr of age were monitored with invasive blood pressure (IBP). The whiskers show 95% CI. A = 0 to 29 days old; B = 29 days to 1 yr old; C = 1 to 4 yr old; D = 4 to 12 yr old; E = 12 to 18 yr old; NIBP = noninvasive blood pressure.

Because most ETCO2 artifacts were related to mask ventilation (table 6), we related anesthetic technique to the occurrence of ETCO2 artifacts and deviating values in a post hoc analysis. ETCO2 artifacts were more common with an inhalation compared to an intravenous induction ($P < 0.001$; table 7). Anesthesia maintenance type was associated with ETCO2 artifact and deviating value incidences (both $P < 0.001$). Among anesthesia maintenance types, the ETCO2 artifact incidence was highest with mask ventilation (28.1% vs. 2.5 to 8.0% with other maintenance types). Children up to 4 yr of age received fewer intravenous inductions compared to older children (0 to 25.9% vs. 42.3 to 71.4%; $P < 0.001$).

Table 7: Anesthetic Technique Related to End-tidal Carbon Dioxide Artifacts and Deviating Values

		Number (%; 95% CI) of ETCO2 artifacts	Number (%; 95% CI) of ETCO2 deviating values
Anesthesia induction type	Inhalation (sevoflurane)	547 (9.0; 8.3-9.7)*	2795 (46.0; 44.7-47.2)
	Intravenous (propofol)	178 (4.9; 4.2-5.6)*	1160 (31.9; 30.4-33.4)
Anesthesia maintenance type	Inhalation (sevoflurane) with tube	473 (8.0; 7.4-8.8) †	2044 (34.8; 33.5-36.0) †
	Inhalation (sevoflurane) with LMA	147 (5.7; 4.8-6.6) †	1645 (63.5; 61.6-65.3) †
	Inhalation (sevoflurane) with mask	81 (28.1; 23.0-33.7) †	148 (51.4; 45.5-57.3) †
	Intravenous (propofol)	24 (2.5; 1.6-3.7) †	118 (12.3; 10.3-14.6) †

*Statistical significance (P < 0.001) between induction types among ETCO2 artifacts.
†Statistical significance (P < 0.001) between maintenance types, among ETCO2 artifacts, and among ETCO2 deviating values.ETCO2 = end-tidal carbon dioxide; LMA = laryngeal mask airway.

Discussion

Each hospital has its unique amount and characteristics of artifacts in its AIMS, which are dependent on the interaction of many factors. The present prospectively studied pediatric population undergoing general anesthesia, with manual artifact data collection live in the operating room, showed that artifacts are present in a substantial number of recordings and that they are associated with deviating status of the measurement, phase of anesthesia, and anesthetic technique.

In addition, a previous study of our group showed that the number and type of artifacts are also dependent on medical specialty and surgical procedure.[15] Moreover, there are many other factors that may influence the number and type of artifacts that include clinical practice (e.g., caseload, standard location of the pulse oximeter probe, timing of venous cannula placement, use of an induction room, type of induction and maintenance of anesthesia, delineation of anesthesia phases), workflow (e.g., human errors), specifications, built-in and selected settings and filters of the patient monitor (e.g., the algorithm for calculation of

the measurements), filters, sampling rate, and the possibility to edit data in the AIMS, among others.[28–30]

The measurement algorithms, filters, and system settings of the various monitors and data storage in AIMS have a large influence on the results of the measurements and the incidence of artifacts. The measurement algorithms for saturation and blood pressure measurement and detection of HR vary among available commercial systems and are known to influence the measurement results and the incidence of artifacts.[28–30] Publication of these algorithms of all commercial systems is essential for clinical practice and research.

The incidence of artifacts in our setting (varying between 0.5% for HR and 7.5% for ETCO2) cannot directly be generalized to other (pediatric) anesthesia settings. However, the concept that electronic data have limited validity is applicable to all systems and hospitals. The limited validity of the data in AIMS should be taken into account not only for research but also when using these data for clinical practice and medicolegal litigations.[1,7–14,31]

Factors Contributing to Artifacts

Artifacts do not occur at random in AIMS data. Vital parameter values deviating from a predefined reference range were more often artifacts than values in a normal range. In addition, the artifact incidence differed among the three anesthesia phases. In general, the artifact incidence (except for noninvasive blood pressure) was higher during the induction and emergence, compared to the maintenance phase (table 4). During the induction and emergence phases there was more patient movement, contributing to a higher artifact incidence. Most artifacts in ETCO2 were related to mask ventilation, which also occurs more often during the induction and emergence phases.

The anesthetic technique influenced the occurrence of ETCO2 artifacts (tables 6 and 7). ETCO2 artifacts were more common during an inhalation compared to an intravenous induction. Because younger children more frequently received an inhalation induction compared to older children, the ETCO2 artifact incidence was also higher in younger age groups. This is also related to the type of airway device. The incidence of ETCO2 artifacts during anesthesia maintenance was highest with mask ventilation and in neonates. With mask ventilation, the dead

space is larger compared to ventilation through an endotracheal tube or laryngeal mask airway.[32] Furthermore, the gradient between ETCO2 and arterial carbon dioxide depends on the dead space–to–tidal volume ratio, which is increased at younger age.[23,32] Also, because tube size is more difficult to estimate in younger children, transient gas leakage due to inadequate sealing before replacing the tube by a larger one contributed to a higher incidence of ETCO2 artifacts. It is known that ETCO2 values are more reliable when cuffed instead of uncuffed tubes are used in this young age group.[33] Although we routinely use cuffed tubes in younger children (data not presented), ETCO2 artifacts due to inadequate sealing occurred in children who were (initially) intubated with an uncuffed tube.

A remarkable number of artifacts appear due to errors on the part of the anesthesiologist, e.g., with the patient not being connected to anesthesia equipment, a wrong-sized blood pressure cuff, or an inappropriately placed invasive blood pressure transducer (table 6). A considerable percentage of these errors are inherent to pediatric anesthesia with a delay in connection to anesthesia equipment in the operating room and difficulties with choosing an appropriate cuff size or are related to the transfer of patients between operating table and bed. These human errors could easily be avoided by the anesthesia staff, by paying attention, and by giving priority to elimination of these errors.

Consequences for Research, Medicolegal Issues, and Clinical Practice

Artifacts represented a substantial proportion of deviating values, which results in erroneous documentation of deviating vital sign values in the AIMS. Artifacts in AIMS data obscure the patient's true physiologic state and confound the interpretation of abnormal vital parameter values. Knowledge of the incidence and causes of artifacts is important for clinical, research, and medicolegal issues.[1,7–14,31] Artifacts can cause an overclassification of deviations when artifacts are within the deviating range, whereas the actual true patient's vitals are within the reference range, or an underclassification of deviations, when artifacts are within the reference range, but the true values are in the deviating range. Implications of these registered artifact values are expected to be of greater importance for database research and litigation procedures than for clinical practice. In clinical practice, it is possible to recognize and to neglect artifacts in the AIMS in the context of all available information, including the graphical representation of vital parameters on the computer screen and the combination of variables and/or

outlying values. On indication, the anesthesia staff can make free text annotations in the AIMS during anesthesia clarifying artifacts, which could be processed to use in database research.

Current artifact filtering methods, which detect outliers or correct for outliers, for example by excluding vital parameter values above and below certain thresholds using an extreme value filter,[34] do not sufficiently deal with artifacts. The artifact filtering method in our AIMS, which corrects for outliers by storing the median or highest value/min, is not perfect either. However, by preventing the influence of short-lasting artifacts by using this low-pass filter, we believe that the values in our AIMS are more valid compared to a method of storing a snapshot or mean value.[11,15]

A possible solution toward a better identification of artifacts in AIMS is an option to allow users to manually (in)validate, remove or change values, or mark values as artifacts.[35] Our AIMS does not have this feature, but even with this possibility, artifacts can still be present in the AIMS database. Because manually produced anesthetic records have shown to be less accurate and complete than computer-generated records,[2–7] one could argue that manual annotations, validations, and markings of artifacts may also be inaccurate and incomplete, because both are human work.

Another option might be the automatic identification of (potential) artifacts by intelligent filtering during data capturing from the monitoring system. This method can be used to filter out artifacts automatically or for a pop-up that questions whether the value is an artifact, every time a value outside a predefined reference range is identified, which must be answered.[31]

The ultimate solution would be an AIMS that automatically filters artifacts from true values using an artifact recognition algorithm,[36] which might be possible by considering more parameters around the time at which the vital sign is measured. The latter option would be promising, because it could be applied retrospectively as well. Another advantage of an algorithm considering medicolegal issues is that it is more objective compared to manual annotations in the AIMS.

Limitations

Despite our efforts to classify artifacts objectively, interobserver differences could have influenced our findings, because anesthesiologists could have given different reasons for similar causes of remarkable, outlying, or deviating values. Labeling as artifacts values that are in fact true values, or vice versa, would result in overestimation or underestimation, respectively, of artifact incidences.

In addition, awareness of anesthesiologists of situations that increase the risk of artifacts and deviating values may have been raised by the presence of the investigator in the operating room and by knowledge of the purposes of this study (Hawthorne effect). This could have resulted in behavior avoiding artifacts and deviations and an underestimation of incidences of artifacts and deviations.[3,37]

Finally, our AIMS documents one single HR value, depending on availability derived from either the electrocardiogram or plethysmograph. It might be possible that the incidence of HR artifacts may be different if each monitoring modality was recorded separately.

Conclusions

The present study showed that vital parameter data in the AIMS are not always valid and hold a considerable number of artifacts. Deviating values are more often artifacts than values in a normal range. Importantly, artifacts do not appear at random in the data, and they are associated with the phase of anesthesia and anesthetic technique. HR and SpO2 have a low incidence of artifacts, whereas mean noninvasive and invasive blood pressure and ETCO2 have a higher incidence of artifacts in our pediatric anesthesia practice. Due to differences between hospitals in clinical practice, workflow, monitoring systems, and AIMSs, our findings cannot directly be generalized to other (pediatric) anesthesia settings. However, the concept that electronic data have limited validity is applicable to all systems and hospitals. The present study highlights the awareness for the presence of artifacts in AIMS data of vital parameters and should stimulate researchers in the anesthesia field to take into account artifacts in their research databases. Furthermore, development of (automatic) data validation systems or solutions to deal with artifacts in data is warranted. Independently of the used equipment and procedures, attention for factors associated with artifacts may lead to ways

to reduce (the influence of) these factors and eventually a reduction of artifacts in AIMS databases.

Research Support

Support was provided solely from institutional and/or departmental sources.

Competing Interests

The authors declare no competing interests.

References

1. Stol, IS, Ehrenfeld, JM, Epstein, RH . Technology diffusion of anesthesia information management systems into academic anesthesia departments in the United States. Anesth Analg 2014; 118:644–50
2. Choi, CK, Saberito, D, Tyagaraj, C, Tyagaraj, K . Organizational performance and regulatory compliance as measured by clinical pertinence indicators before and after implementation of Anesthesia Information Management System (AIMS). J Med Syst 2014; 38:5
3. Edwards, KE, Hagen, SM, Hannam, J, Kruger, C, Yu, R, Merry, AF . A randomized comparison between records made with an anesthesia information management system and by hand, and evaluation of the Hawthorne effect. Can J Anaesth 2013; 60:990–7
4. Eden, A, Grach, M, Goldik, Z, Shnaider, I, Lazarovici, H, Barnett-Griness, O, Perel, A, Pizov, R . The implementation of an anesthesia information management system. Eur J Anaesthesiol 2006; 23:882–9
5. Reich, DL, Wood, RKJr, Mattar, R, Krol, M, Adams, DC, Hossain, S, Bodian, CA . Arterial blood pressure and heart rate discrepancies between handwritten and computerized anesthesia records. Anesth Analg 2000; 91:612–6
6. Edsall, DW, Deshane, P, Giles, C, Dick, D, Sloan, B, Farrow, J . Computerized patient anesthesia records: Less time and better quality than manually produced anesthesia records. J Clin Anesth 1993; 5:275–83
7. Anderson, BJ, Merry, AF . Paperless anesthesia: Uses and abuses of these data. Paediatr Anaesth 2015; 25:1184–92
8. Muravchick, S, Caldwell, JE, Epstein, RH, Galati, M, Levy, WJ, O'Reilly, M, Plagenhoef, JS, Rehman, M, Reich, DL, Vigoda, MM . Anesthesia information management system implementation: A practical guide. Anesth Analg 2008; 107:1598–608
9. Van Cleve, WC, Nair, BG, Rooke, GA . Associations between age and dosing of volatile anesthetics in 2 academic hospitals. Anesth Analg 2015; 121:645–51
10. Wanderer, JP, Anderson-Dam, J, Levine, W, Bittner, EA . Development and validation of an intraoperative predictive model for unplanned postoperative intensive care. Anesthesiology 2013; 119:516–24
11. de Graaff, JC, Bijker, JB, Kappen, TH, van Wolfswinkel, L, Zuithoff, NP, Kalkman, CJ . Incidence of intraoperative hypoxemia in children in relation to age. Anesth Analg 2013; 117:169–75
12. Bell, G . Lessons for pediatric anesthesia from audit and incident reporting. Paediatr Anaesth 2011; 21:758–64

13. Vigoda, MM, Lubarsky, DA . The medicolegal importance of enhancing timeliness of documentation when using an anesthesia information system and the response to automated feedback in an academic practice. Anesth Analg 2006; 103:131–6

14. Feldman, JM . Do anesthesia information systems increase malpractice exposure? Results of a survey. Anesth Analg 2004; 99:840–3

15. Kool, NP, van Waes, JA, Bijker, JB, Peelen, LM, van Wolfswinkel, L, de Graaff, JC, van Klei, WA . Artifacts in research data obtained from an anesthesia information and management system. Can J Anaesth 2012; 59:833–41

16. Jubran, A . Pulse oximetry. Crit Care 2015; 19:272

17. Butani, L, Morgenstern, BZ . Are pitfalls of oxcillometric blood pressure measurements preventable in children? Pediatr Nephrol 2003; 18:313–8

18. Spiegel, J. End-tidal carbon dioxide: The most vital of vital signs. Anesthesiol News 2013; 21–7October:

19. de Graaff, JC, Sarfo, MC, van Wolfswinkel, L, van der Werff, DB, Schouten, AN . Anesthesia-related critical incidents in the perioperative period in children: A proposal for an anesthesia-related reporting system for critical incidents in children. Paediatr Anaesth 2015; 25:621–9

20. Moller, JT, Johannessen, NW, Berg, H, Espersen, K, Larsen, LE . Hypoxaemia during anaesthesia: An observer study. Br J Anaesth 1991; 66:437–44

21. Thorpe, CM, Gauntlett, IS . Arterial oxygen saturation during induction of anaesthesia. Anaesthesia 1990; 45:1012–5

22. Gravenstein, JS . Gravenstein, JS, Jaffe, MB, Gravenstein, N, Paulus, DA . Clinical perspectives, Capnography, 2nd edition. Edited by 2011, pp New York, Cambridge University Press, 1–8

23. McSwain, SD, Hamel, DS, Smith, PB, Gentile, MA, Srinivasan, S, Meliones, JN, Cheifetz, IM . End-tidal and arterial carbon dioxide measurements correlate across all levels of physiologic dead space. Respir Care 2010; 55:288–93

24. Sullivan, KJ, Kissoon, N, Goodwin, SR . End-tidal carbon dioxide monitoring in pediatric emergencies. Pediatr Emerg Care 2005; 21:327–35

25. Bonafide, CP, Brady, PW, Keren, R, Conway, PH, Marsolo, K, Daymont, C . Development of heart and respiratory rate percentile curves for hospitalized children. Pediatrics 2013; 131:e1150–7

26. de Graaff, JC, Pasma, W, van Buuren, S, Duijghuisen, JJ, Nafiu, OO, Kheterpal, S, van Klei, WA . Reference values for noninvasive blood pressure in children during anesthesia: A multicentered retrospective observational cohort study. Anesthesiology 2016; 125:904–13

27. Arts, DG, De Keizer, NF, Scheffer, GJ . Defining and improving data quality in medical registries: A literature review, case study, and generic framework. J Am Med Inform Assoc 2002; 9:600–11

28. Kaufmann, MA, Pargger, H, Drop, LJ . Oscillometric blood pressure measurements by different devices are not interchangeable. Anesth Analg 1996; 82:377–81

29. Petterson, MT, Begnoche, VL, Graybeal, JM . The effect of motion on pulse oximetry and its clinical significance. Anesth Analg 2007; 105(6 suppl):S78–84

30. Hay, WWJr, Rodden, DJ, Collins, SM, Melara, DL, Hale, KA, Fashaw, LM . Reliability of conventional and new pulse oximetry in neonatal patients. J Perinatol 2002; 22:360–6

31. Nair, BG, Horibe, M, Newman, SF, Wu, WY, Peterson, GN, Schwid, HA . Anesthesia information management system-based near real-time decision support to manage intraoperative hypotension and hypertension. Anesth Analg 2014; 118:206–14

32. Casati, A, Fanelli, G, Torri, G . Physiological dead space/tidal volume ratio during face mask, laryngeal mask, and cuffed oropharyngeal airway spontaneous ventilation. J Clin Anesth 1998; 10:652–5

33. Weiss, M, Dullenkopf, A, Fischer, JE, Keller, C, Gerber, AC ; European Paediatric Endotracheal Intubation Study Group: Prospective randomized controlled multi-centre trial of cuffed or uncuffed endotracheal tubes in small children. Br J Anaesth 2009; 103:867–73

34. Sottas, CE, Cumin, D, Anderson, BJ . Blood pressure and heart rates in neonates and preschool children: An analysis from 10 years of electronic recording. Paediatr Anaesth 2016; 26:1064–70

35. Wax, DB, Beilin, Y, Hossain, S, Lin, HM, Reich, DL . Manual editing of automatically recorded data in an anesthesia information management system. Anesthesiology 2008; 109:811–5

36. Gostt, RK, Rathbone, GD, Tucker, AP . Real-time pulse oximetry artifact annotation on computerized anaesthetic records. J Clin Monit Comput 2002; 17:249–57

37. Rheineck-Leyssius, AT, Kalkman, CJ, Trouwborst, A . Influence of motivation of care providers on the incidence of postoperative hypoxaemia in the recovery room. Br J Anaesth 1996; 77:453–7

5

Artifact processing methods influence on intraoperative hypotension quantification and outcome effect estimates

Wietze Pasma
Linda M. Peelen
Stef van Buuren
Wilton A. van Klei
Jurgen C. de Graaff

Published in Anesthesiology 2020 Apr;132(4):723-737

Abstract

Background

Physiologic data that is automatically collected during anesthesia is widely used for medical record keeping and clinical research. These data contain artifacts, which are not relevant in clinical care, but may influence research results. The aim of this study was to explore the effect of different methods of filtering and processing artifacts in anesthesiology data on study findings in order to demonstrate the importance of proper artifact filtering.

Methods

The authors performed a systematic literature search to identify artifact filtering methods. Subsequently, these methods were applied to the data of anesthesia procedures with invasive blood pressure monitoring. Different hypotension measures were calculated (i.e., presence, duration, maximum deviation below threshold, and area under threshold) across different definitions (i.e., thresholds for mean arterial pressure of 50, 60, 65, 70 mmHg). These were then used to estimate the association with postoperative myocardial injury.

Results

After screening 3,585 papers, the authors included 38 papers that reported artifact filtering methods. The authors applied eight of these methods to the data of 2,988 anesthesia procedures. The occurrence of hypotension (defined with a threshold of 50 mmHg) varied from 24% with a median filter of seven measurements to 55% without an artifact filtering method, and between 76 and 90% with a threshold of 65 mmHg. Standardized odds ratios for presence of hypotension ranged from 1.16 (95% CI, 1.07 to 1.26) to 1.24 (1.14 to 1.34) when hypotension was defined with a threshold of 50 mmHg. Similar variations in standardized odds ratios were found when applying methods to other hypotension measures and definitions.

Conclusions

The method of artifact filtering can have substantial effects on estimates of hypotension prevalence. The effect on the association between intraoperative hypotension and postoperative myocardial injury was relatively small. Nevertheless, the authors recommend that researchers carefully consider artifacts handling and report the methodology used.

Introduction

Physiologic data that is automatically collected are widely used for medical record keeping and clinical research; yet, not every stored vital sign in these data represents the actual physiologic state of the patient at the time of measurement. For example, when a person leans against a patient's blood pressure cuff, this causes artifacts in blood pressure data, i.e., the value that is registered does not equal the true blood pressure of the patient. Similarly, electrocautery disturbs electrocardiogram readings and causes artifacts in heart rate data. These artifacts are not relevant in clinical care, as they are easily recognized and subsequently ignored, but may influence research results.[1,2] Consequently, before using the physiologic data for research, one should consider how to process artifacts, and this should be reported when presenting the results of the study.[1,3–5]

To make artifactual data more suitable for retrospective analyses, the researcher has several options, varying from manual data cleaning via simple filtering methods to more advanced methods. Methods will vary in their ability to identify or correct artifacts in the data.[3,6] More importantly, different filter methods can lead to different study results. Previous studies have shown that artifact occurrence depends on several factors, which could lead to incorrect estimation of associations in clinical research, i.e., a form of misclassification bias.[4,5,7,8] We hypothesized that these filter methods used on the data would have a significant influence on the findings of the study.

The aim of this study was to explore the effect of different methods of filtering and processing artifacts in anesthesiology data on study findings. In doing so, we aimed to demonstrate the importance of proper artifact filtering. To this aim, we performed a systematic literature search to identify artifact filtering methods actually being used in practice, and subsequently used these filters on an existing perioperative dataset. For this dataset, we selected surgical patients in whom both invasive blood pressure measurements and an example outcome (myocardial injury) were measured. We examined the effects of artifact filtering methods on the quantification of intraoperative hypotension measures and subsequently its effect on the association between intraoperative hypotension and postoperative myocardial injury. The determinant invasive blood pressure measurement was chosen because it has a known high incidence of artifacts and is often used in intraoperative hypotension research.[2,5]

Methods

Systematic search for artifact filtering methods

To find relevant artifact filtering methods, we performed a systematic literature search. No formal review protocol was developed before beginning the review. We started the search with the following search query on PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) on June 8, 2018:

```
(anesthesia [tiab] OR aims [tiab] OR intraoperative [tiab] OR "Monitoring, Intraoperative"[Mesh] OR "Anesthesia"[Mesh] OR "Monitoring, Physiologic"[Mesh] OR "Heart Rate"[Mesh] OR "Oximetry"[Mesh] OR "Blood Pressure"[Mesh] OR "Arterial Pressure"[Mesh] OR "blood pressure" [tiab] OR "arterial pressure" [tiab] OR "oxygen saturation" [tiab] OR "heart rate" [tiab]) AND (artefact [tiab] OR artifact [tiab] OR "Artifacts"[Mesh] OR "measurement error" [tiab]).
```

First, two researchers (W.P. and L.P.) screened every title for possible inclusion. Exclusion reasons were noted and categorized. Conflicts between the reasons of exclusion were reviewed and the first reviewer made a definitive choice based on the title or abstract. Papers were included when: (1) it was published after 2000; (2) it used vital signs data for anesthesia (oxygen saturation, blood pressure, or heartrate); and (3) a process of artifact filtering was described. Papers about the development of a filtering method were included, as well as research papers in which a specific filtering method was applied and review papers that discussed different methods. We excluded case reports, letters to editors, editorials, and studies that considered nonhuman subjects. Apart from these, no other limits were placed on the type of study design. Papers in languages other than English were also not considered. We only considered papers that came up with our search query, we did not review reference sections of papers to identify additional candidates. No effort was made to consider unpublished studies, conference abstracts, or proceedings.

After title screening, all papers selected by one or both of the reviewers underwent abstract screening and, if necessary, fulltext screening using the aforementioned approach. Abstracts were extracted from PubMed and displayed in a review form (programed in R shiny), accompanied by the title and a link to the manuscript. The

final decision and exclusion reason were filled in using dropdown fields. The form also included the choices that both reviewers made regarding title screening. If it was unclear from the abstract text, if any artifact filtering methods were described in the paper, the manuscript was screened to make this decision. From the remaining papers, any method for artifact filtering was collected and categorized. The first author screened the abstracts and identified and classified the artifact filtering methods within the selected papers. The three artifact filtering method categories described in the results section were identified and refined during the review process. Some papers included multiple types or a combination of artifact filtering methods. Therefore, one paper could fall under several artifact filtering methods. We did not contact authors of original reports to clarify the methods that they used. The information in the included papers was sufficient to categorize the methods.

Cohort

To evaluate the influence of the obtained methods of artifact filtering on the intraoperative hypotension measures, we selected a group of patients from a prospectively-defined and previously-described cohort.[9] In short, in this ongoing cohort patients are included if they are 60 yr of age or older and undergo intermediate- to high-risk noncardiac surgical procedures at the University Medical Center Utrecht. We selected only new procedures, i.e., reoperations were not included. A first procedure for a patient was defined as a procedure with an available troponin measurement during the first 3 postoperative days and not preceded in the previous 365 days by another eligible procedure. We further refined the cohort by selecting patients in whom sufficient postinduction invasive blood pressure measurements were available, i.e., we excluded procedures without a known time of induction and without at least 15 invasive blood pressure measurements after the start of induction throughout the entire procedure. For the current study, procedures from January 1, 2011 to December 31, 2014 were selected. The patient cohort selection and further analysis described in the current study were designed for illustration purposes only. For collection of these data, the local ethics committees approved the protocol and waived the need for informed consent (University Medical Center Utrecht Medical Research Ethics Committee, protocol no. 18-261).

We collected invasive mean blood pressure measurements from our anesthesia information management system (Anstat, version 2.0.4, 2015; Carepoint, Ede, The Netherlands). Invasive blood pressure is measured during anesthesia with an IntelliVue monitoring system (type MP70, X2 multimeasurement module; Philips, Germany) with a built-in filter for artifacts (i.e., a 12-Hz filter is applied) that reduces resonant effects of the tubing system. Our anesthesia information management system stores one value each minute in its database, which is the median of 12 consecutive measurements with a 5-s interval supplied by the anesthesia monitor. We determined the time of induction with an algorithm that was previously published.[10] Blood pressure measurements before time of induction were excluded from analysis.

Artifact filtering methods

From the artifact filtering methods resulting from the systematic search, we included methods into our comparison based on their frequency of use (frequently used methods were more likely to be included), the need for annotated data for the purpose of algorithm training, and whether they are applicable to anesthesia information management system data (i.e., minute to minute vital sign data). In addition, as many of the artifact filtering methods contain cutoff values, we varied such cut-off values to assess differences within the same type of filter. For example, if filters are based on distribution, we used "more than [two] times the interquartile range" and "more than [three] times the interquartile range." A detailed description of the filtering methods compared can be found in the Results section, after the results of the systematic search.

Outcome

The primary outcome was the incidence and severity of intraoperative hypotension when applying different artifact filtering methods found in the review step. We considered four measures to quantify hypotension; the presence of hypotension, the total duration of hypotension, the total area under threshold, and the maximum deviation below the threshold if hypotension was present. Hypotension was quantified by an algorithm, which took every measurement and made a linear interpolation between every data point. The interpolation was performed between subsequent measurements, regardless of the period between both measurements. The area between the threshold and this blood pressure curve was then quantified.[11,12] Each time the blood pressure curve dropped under a specified threshold,

this was identified as the start of an hypotension episode, and as soon as the blood pressure went over the threshold again, this point in time was marked as the end of that episode. The total area under the threshold is the summation of all episodes. We used four different thresholds (mean blood pressure of 50, 60, 65, and 70 mmHg) to explore whether artifact filters can have a different effect under different hypotension definitions.[13] Altogether we compared the artifact filtering methods for four different hypotension measures and four different thresholds.

The secondary outcome was the association between hypotension and postoperative myocardial injury, when applying different artifact methods. Postoperative myocardial injury was defined as a postoperative elevation of troponin I. According to local protocol, troponin was measured routinely for the first 3 postoperative days. Troponin elevation was present when at least one of the postoperative troponin values was more than a predefined clinical cut-off value of greater than 60 ng/l.[9]

Statistical analysis

For each of the artifact filtering methods, we first describe the incidence or extent of hypotension that would result from applying the method to the dataset. We describe intraoperative hypotension as a proportion, and use median values with interquartile ranges for the continuous hypotension measures (total area under threshold, total duration, and maximum deviation from threshold). This is done for each of the thresholds for intraoperative hypotension.

To allow for a comparison of the effect estimate in the association between hypotension and postoperative myocardial injury, we standardized hypotension measures per measure type and threshold before estimating the effect of hypotension on postoperative myocardial injury. In order to do so, we first used a logarithmic transformation for total duration and total area under threshold, since these data were skewed to the right. After this transformation, the hypotension measures were standardized by calculating a z-score. With these standardized hypotension measures as the only explaining variable, we fitted a model for postoperative myocardial injury using logistic regression analysis. This was done again for each combination of artifact filtering method, hypotension measure, and hypotension threshold. We expressed effect estimates of hypotension measures as standardized odds ratios with CIs, using a level of significance of α=0.05.

We collected and deidentified the data from the enterprise data warehouse with SAS software (Version 9.4, SAS Institute Inc). After deidentification, the data was further processed and analyzed in R (R Foundation for Statistical Computing, Vienna, Austria; https://www.R-project.org, R version 3.5.1 [2018–07–02]).

Results

Systematic search for artifact filtering methods

The systematic search for artifact filtering methods resulted in 3,585 papers. After abstract and full-text screening, 3,300 papers were excluded and 285 papers mentioning artifact filtering methods remained - of which 247 papers described methods used on high resolution data such as electrocardiogram, photoplethysmogram, and arterial blood pressure waveform data (figure 1). These methods often rely on the repeatable patterns in these signals, and therefore are not applicable to anesthesia information management system data which is used for clinical research. In addition, a number of filters on these high-density data rely on additional sensor data such as accelerometer data, which are not commonly collected in an anesthesia information management system database. After excluding this group, 38 papers remained that report methods applicable to 1-min resolution anesthesia information management system data (figure 1).

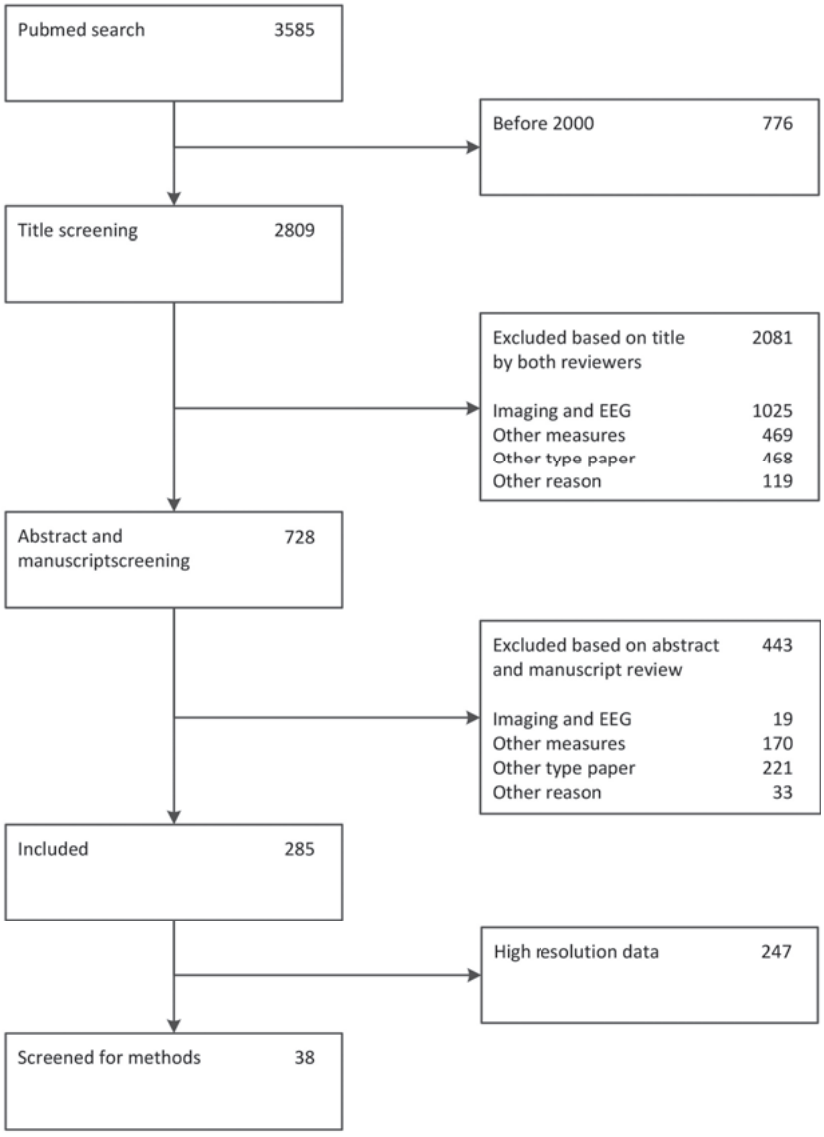


Figure 1: Literature review flowchart. Flow diagram illustrating systematic search and review process, starting with 3,585 papers. Eventually 38 papers remained for identification of artifact filtering methods. Papers about imaging included papers of magnetic resonance imaging, computed tomography echo, and other imaging techniques. Papers about other measures include nonstandard vital sign monitors such as camera-, textile-, or wristwatch-based monitors. Other types of papers were methodologic papers or original research in which physiological data processing was not used or mentioned. Other reasons for exclusion were for example papers on pathology or ophthalmology. EEG, electroencephalography.

We divided the artifact filtering methods into three basic categories. The first category consisted of methods that identify measurements as an artifact, such as limit methods that use biologic plausible blood pressure boundaries (mentioned in 27 papers). The second group consisted of methods that can alter the vital sign signal by applying a filter to extract the true signal of interest from the raw artefactual data (mentioned in 16 papers), for example a method that calculates the median over neighboring values. This new acquired data point is then used to calculate hypotension. The third category contained methods that use model strategies that take artifacts or measurement errors into account (mentioned in six papers). This can be done by applying a model to the data, such as a spline function of the blood pressure, which is subsequently fit into a model instead of using the actual data. All 38 papers included, and the methodologies found in these papers are listed in figure 2.

Based on the results of the systematic search we chose three methods to handle artifacts in invasive blood pressure data: a limit filter, a moving median filter, and a likelihood filter based on median and interquartile range. For each of the methods we used two or three different settings for the parameters. Altogether, we compared eight different filtering approaches (including applying no artifact filtering method), which are described in detail in figure 3.

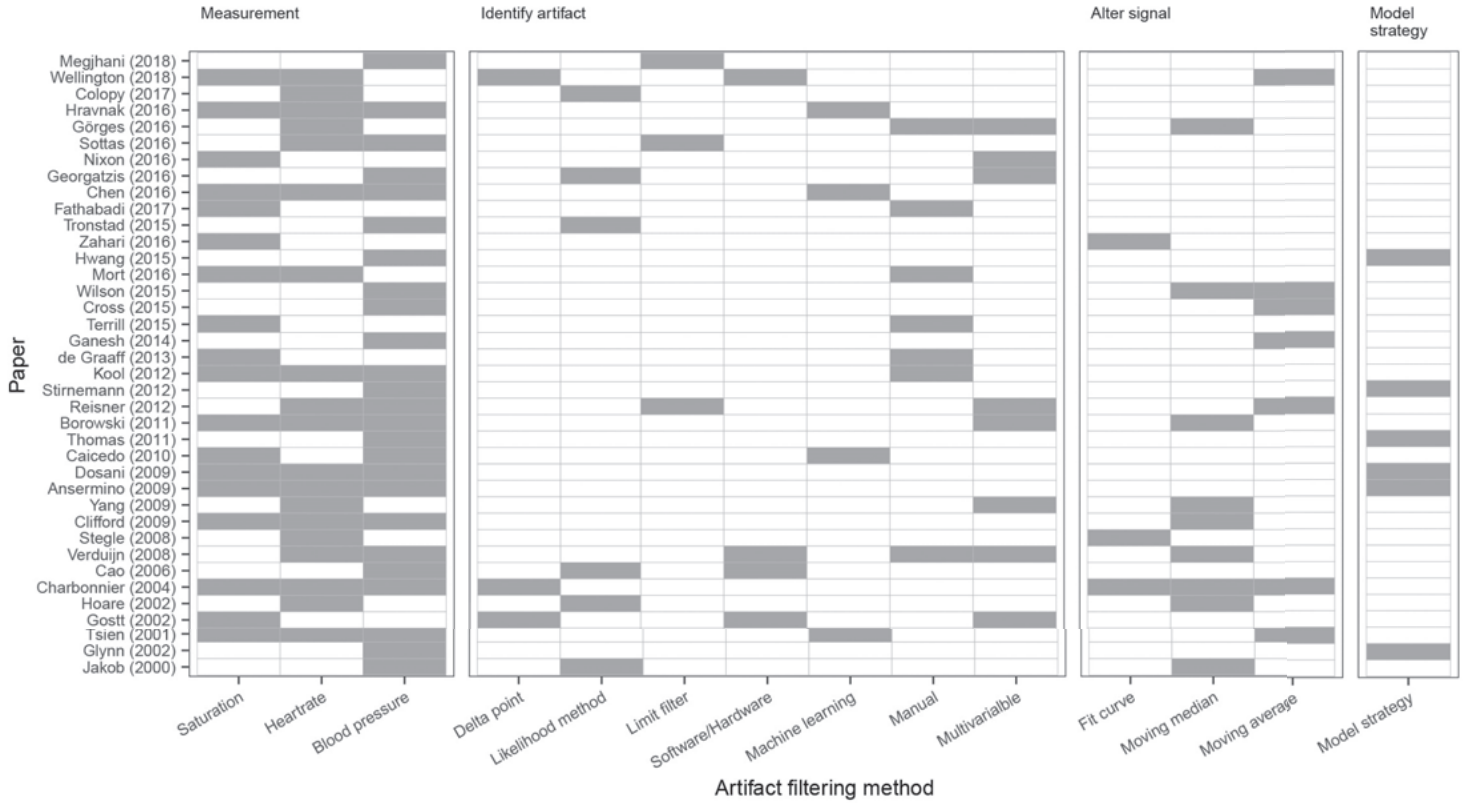


Figure 2: Literature overview artifact methods. Methods within papers (vertical) found with systematic search (38 papers). The first panel describes which measurements are described in the paper. The other three panels represent the method groups that are described in this paper. Each method (horizontal) that was described in the paper was marked. One paper can contain more than one method. [2,3,15–27,30–53]

First, **no filter** was applied, i.e., all measurements after time of induction were used to calculate hypotension.

Second, we used a **limits filter** by excluding invasive blood pressure readings with a corresponding pulse pressure lower than 20 or higher than 150. In addition, the mean blood pressure had to be between 40 and 160 mmHg.*

Third, we applied a **moving median** to the blood pressure curve that replaced every blood pressure with the median value of n readings, n being the window size, of which the current reading is the center measurement.†

$$x'_i = med[x_{i-((n-1)/2)}, \dots, x_i, \dots, x_{i+((n-1)/2)}]$$

Finally we applied a **likelihood filter** based on a median and variation within a time window. We divided the blood pressure signal into subsequent windows of 10 measurements. For each window the median and IQR were calculated. If a blood pressure deviated more than n × IQR, but at least 10 mmHg from the median, the value was identified as an artifact and was removed from the signal.‡

Figure 3: List of artifact filtering methods applied to blood pressure data. All blood pressure measurements were used to quantify hypotension. Four artifact filtering methods were chosen, which were applied on blood pressure data. *As a variation, the lower limits for pulse pressure and mean blood pressure were varied, to 10 mmHg and 20 mmHg, respectively.[17–19] †Two different values for n (5 and 7) were used.[3,20–27] ‡Three different values for n (1, 2, and 3) were used.[15,16] IQR, interquartile range.

Cohort and hypotension measures

We included 2,988 anesthetic procedures in our analysis (figure 4), of which the baseline characteristics are listed in table 1. In this cohort the occurrence of postoperative myocardial injury was 807 (27%), and 1,563 procedures were classified as high-risk surgeries.

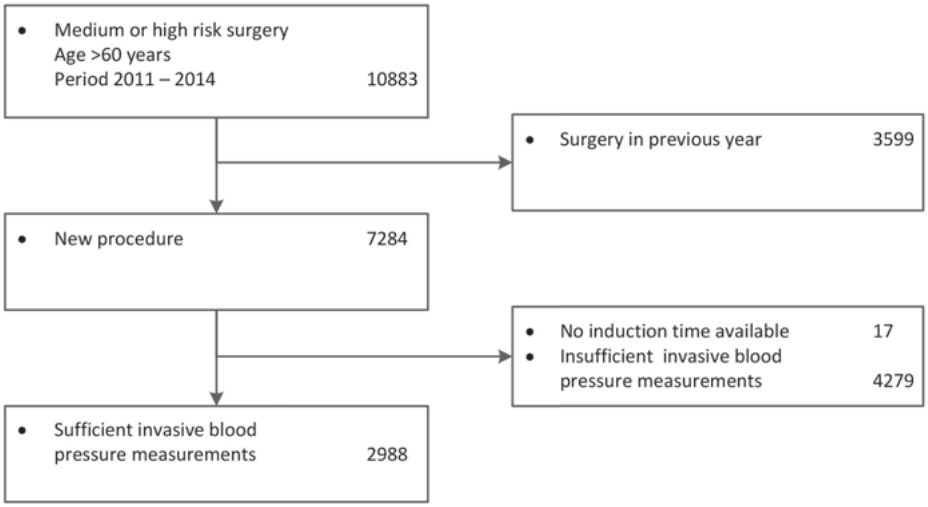


Figure 4: Flowchart procedure inclusion. All procedures identified in the anesthetic information management system, and those included in our analyses. Included procedures were required to be medium- to high-risk procedures between 2011 and 2014 on patients who were 60 yr or older. Procedures should be new (no procedures on the same patient in the previous yr) and have sufficient (at least 15) invasive blood pressure measurements available after time of induction.

Table 1: Baseline data included procedures (n = 2,988)

	N or median	Percentage or Interquartile Range
Postoperative myocardial injury (%)	807	(27.0%)
Number of blood pressure measurements [IQR]	203	[143-297]
Age at time of surgery, years [IQR]	70.3	[65.6-76.5]
Male (%)	1,861	(62.3%)
ASA classification (%)		
I	266	(8.9%)
II	1,635	(54.7%)
III	978	(32.7%)
IV-V	109	(3.6%)

Table 1 *(continued)*: Baseline data included procedures (n = 2,988)

	N or median	Percentage or Interquartile Range
Surgical specialty (%)		
General	750	(25.1%)
Orthopedic	142	(4.8%)
Urological	106	(3.5%)
ENT and Dental	287	(9.6%)
Vascular	848	(28.4%)
Neurological	716	(24.0%)
Gynecological	53	(1.8%)
Other	86	(2.9%)
Emergency surgery (%)	579	(19.4%)
High risk surgery (%)	1,563	(52.3%)
Ischemic heart disease (%)	470	(15.7%)
History of cerebrovascular disease (%)	710	(23.8%)
Chronic heart failure (%)	113	(3.8%)
Peripheral vascular disease (%)	446	(14.9%)
Hypertension (%)	1,670	(55.9%)
Use of beta blockers (%)	1,041	(34.8%)
Use of calcium antagonists (%)	587	(19.6%)
Use of ACE inhibitors or AT2 blockers (%)	1,223	(40.9%)
Use of statins (%)	1,345	(45.0%)
Use of insulin/insulin dependency (%)	200	(6.7%)

Data are presented as number of procedures and percentage of total cohort, except age and number of blood pressure measurements, which are presented as median and IQR. ACE, angiotensin-converting enzyme; ASA, American Society of Anesthesiologists; AT2, angiotensin II receptor; ENT, ear, nose, and throat; IQR, interquartile range.

Table 2 describes the estimated values for the different hypotension measures when using different hypotension thresholds for each of the artifact filtering methods. Different artifact filtering methods resulted in different estimates of the occurrence of hypotension. For example, when hypotension was defined as mean blood pressure below 50 mmHg, the occurrence varied from 24% with a moving

median filter of seven measurements to 55% without an artifact filtering method. When a threshold of 65 mmHg was used, the presence of hypotension varied between 76 and 90%. Similarly, other hypotension measures varied among different artifact filtering methods. For example, within the definition of hypotension as blood pressure less than 65 mmHg, the total area under threshold varied between 81 mmHg × min (interquartile range, 3 to 311) and 129 mmHg × min (interquartile range, 25 to 383), the maximum deviation below threshold varied between 8 mmHg (interquartile range, 1 to 15) and 17 mmHg (interquartile range, 9 to 31), and the total duration of hypotension between 18 min (interquartile range, 2 to 52) and 22 min (interquartile range, 6 to 58).

Association between hypotension and POMI

Figure 5 depicts the standardized odds ratios for the association between intraoperative hypotension and postoperative myocardial injury when using different artifact filtering methods for blood pressure. Again, this is shown for different hypotension measures and different thresholds. Standardized odds ratios ranged from 1.16 (95% CI, 1.07 to 1.26) for not using an artifact filter method to 1.24 (95% CI, 1.14 to 1.34) for using a limit filter (pulse pressure greater than 20; mean blood pressure greater than 40) within the 50-mmHg threshold for presence of hypotension. For the 65-mmHg thresholds, estimates ranged from 1.14 (95% CI, 1.04 to 1.26) for using a limit filter (pulse pressure greater than 20; mean blood pressure greater than 40) to 1.21 (95% CI, 1.11 to 1.33) for using a likelihood filter method (1 × IQR above median). Within the 65-mmHg threshold, the odds ratios for maximum deviation ranged from 1.20 (95% CI, 1.10 to 1.31) to 1.30 (95% CI, 1.19 to 1.42); total area under threshold, 1.27 (95% CI, 1.18 to 1.37) to 1.36 (95% CI, 1.24 to 1.49); and total duration, 1.26 (95% CI, 1.17 to 1.37) to 1.32 (95% CI, 1.21 to 1.44).

Table 2: Overview of hypotension measures

Hypotension		No filter	Limits		Moving median		Likelihood		
Measure	**Definition**	**No filter**	**PP > 20 mABP > 40**	**PP >10 mABP > 20**	**Five-measure window**	**Seven-measure window**	**3xIQR**	**2xIQR**	**1xIQR**
Presence of hypotension (yes/no)	50	1654 (55%)	1201 (40%)	1364 (46%)	876 (29%)	705 (24%)	1394 (47%)	1284 (43%)	1058 (35%)
	60	2451 (82%)	2288 (77%)	2340 (79%)	1979 (66%)	1783 (60%)	2337 (78%)	2289 (77%)	2155 (72%)
	65	2699 (90%)	2609 (88%)	2639 (89%)	2401 (80%)	2278 (76%)	2638 (88%)	2614 (87%)	2528 (85%)
	70	2831 (95%)	2776 (93%)	2793 (94%)	2652 (89%)	2581 (86%)	2794 (94%)	2778 (93%)	2738 (92%)
Maximum decrease (mmHg)	50	2 (0 - 16)	0 (0 - 4)	0 (0 - 7)	0 (0 - 2)	0 (0 - 0)	0 (0 - 8)	0 (0 - 6)	0 (0 - 4)
	60	12 (4 - 26)	8 (1 - 14)	9 (2 - 17)	5 (0 - 12)	3 (0 - 10)	10 (2 - 18)	9 (1 - 16)	7 (0 - 14)
	65	17 (9 - 31)	13 (6 - 19)	14 (7 - 22)	10 (3 - 17)	8 (1 - 15)	15 (7 - 23)	14 (6 - 21)	12 (5 - 19)
	70	22 (14 - 36)	18 (11 - 24)	19 (12 - 27)	15 (8 - 22)	13 (6 - 20)	20 (12 - 28)	19 (11 - 26)	17 (10 - 24)
Total area under threshold (mmHg * min)	50	1 (0 - 24)	0 (0 - 4)	0 (0 - 9)	0 (0 - 5)	0 (0 - 0)	0 (0 - 14)	0 (0 - 11)	0 (0 - 5)
	60	43 (4 - 158)	25 (0 - 104)	30 (1 - 119)	21 (0 - 121)	11 (0 - 103)	36 (1 - 146)	33 (0 - 141)	23 (0 - 124)
	65	128 (25 - 383)	96 (16 - 291)	104 (18 - 316)	95 (10 - 344)	81 (3 - 311)	117 (20 - 370)	114 (18 - 364)	97 (11 - 342)
	70	326 (86 - 805)	266 (68 - 705)	276 (72 - 727)	279 (58 - 761)	259 (44 - 734)	307 (76 - 794)	298 (72 - 785)	284 (62 - 765)
Total duration (min)	50	0(0 - 4)	0(0 - 2)	0(0 - 3)	0(0 - 2)	0(0 - 0)	**0**(0 - 4)	0(0 - 3)	0(0 - 3)
	60	8(1 - 25)	7(0 - 23)	7(1 - 23)	6(0 - 22)	5(0 - 20)	**8**(1 - 25)	8(1 - 25)	7(0 - 24)
	65	22(6 - 58)	20(5 - 55)	21(5 - 57)	19(4 - 54)	18(2 - 52)	**22**(5 - 57)	21(5 - 57)	21(4 - 57)
	70	48(14 - 110)	47(13 - 109)	47(14 - 110)	45(12 - 107)	44(10 - 106)	**48**(14 - 110)	48(14 - 110)	47(13 - 110)

Intraoperative hypotension measures after application of different artifact filtering methods. Continuous data (maximum decrease below threshold, total AUT, and total duration) are presented as median with IQR, and binomial data (presence of hypotension) are presented as number of procedures and percentage of the total procedures. AUT, area under threshold; IQR, interquartile range; mABP, mean arterial blood pressure; PP, pulse pressure.

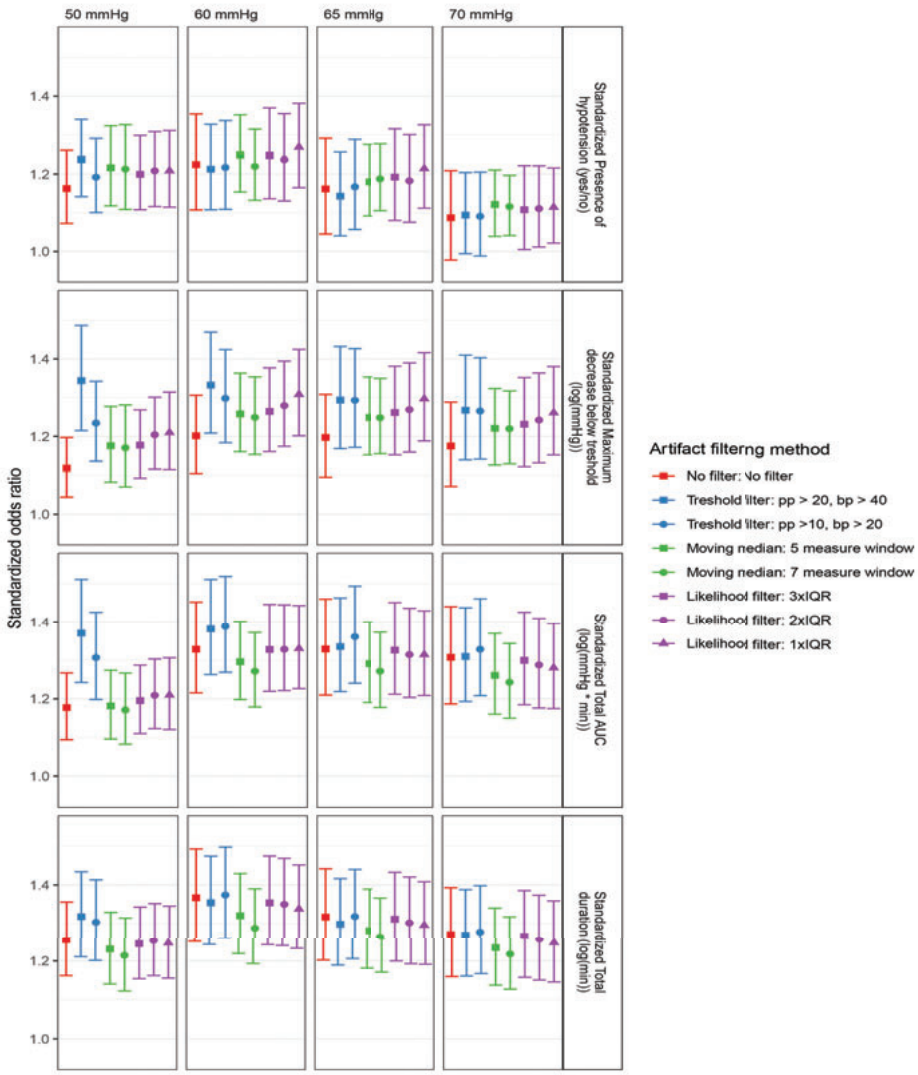


Figure 5: Effect of artifact filtering methods on the association between intraoperative hypotension and postoperative myocardial injury. Standardized odds ratios for the association between intraoperative hypotension and postoperative myocardial injury and for different artifact filtering methods (*color and shapes*). Different hypotension measures (*vertical*) and hypotension definitions (*horizontal*) were compared.

Discussion

Effect of artifact filtering on hypotension measures

Different methods for processing artifacts result in different estimates of hypotension measures. We saw a change in the number of patients who were identified as having intraoperative hypotension when different artifact filtering

methods were applied. The intraoperative hypotension ranged from 24 to 55% when hypotension was defined as mean blood pressure less than 50 mmHg and from 76 to 90% when the defined threshold was 65 mmHg. Although we found this clear effect on hypotension measures, the resulting effect on the association between determinant and outcome (i.e., intraoperative hypotension and postoperative myocardial injury) was less profound than expected.

From previous studies we learned that the occurrence of artifacts in physiologic data are related to patient and procedure characteristics,[2,5] hence we expected changes in estimates when these artifacts were dealt with differently. In the current analyses, removing artifactual data did indeed change the associations, but overall these changes were smaller than the variation in estimates due to the choice of hypotension threshold or choice of hypotension quantity (figure 5). We found filtering methods to have less of an effect on the association between duration and outcome than between depth of hypotension and outcome. Filtering methods for artifacts are designed to correct extreme values or outliers in the data, resulting in adjustments in the depth of hypotension domain rather than the duration of hypotension.

Of note, the odds ratios for hypotension defined according to 50- and 70-mmHg thresholds were low, in contrast to those for 60- and 65-mmHg thresholds. This is explained by the low or high intraoperative hypotension when a low or a high threshold is chosen, respectively. In both situations, the variance of the hypotension measure in the data will be low, compared to choosing an intermediate threshold (i.e., 60 or 65 mmHg). With a higher variance, the distinctive power of the data used to model a given outcome, increases.

We studied the association between intraoperative hypotension and postoperative myocardial injury, because there is an extensive body of clinical research on this topic and on the methodology of analyzing hypotension and outcome due to hypotension. Significant effort has been put into determining a consensus on the hypotension threshold that should be used and on the kind of hypotension measure that should be analyzed. From our analysis we deduce that these choices are probably more important (as they yield more variation) than the choice of artifact filtering method. Nevertheless, we think that artifact filtering methods are an additional source of variance in studying the relationship between intraoperative

hypotension and outcomes, which was already muddled with inconsistencies in methodology.[10,11,13,14]

Systematic search

We performed a systematic search to make sure that the artifact filtering methods we chose corresponded to the manner in which researchers handle artifacts. We divided the methods we found in three categories. The first category includes methods that identify artifacts. These methods include a bandwidth filter with a minimum and maximum allowed value for blood pressure or pulse pressure and a likelihood filter based on the distribution of the data, as used in this paper.[15–19] More advanced methods exist (e.g., models or machine learning algorithms that identify artifacts); although, the downfall of these more advanced methods is that they require a training set to train an algorithm, of which obtaining is labor intensive. When artifacts are identified and subsequently removed, there are several options that can be used to fill in the gaps. In the current paper, we used linear interpolation between existing data points,[12] but other methods are also available, such as last known value carried forward.

The second category is comprised of methods that replace the data signal with a new and refined signal. An example of these methods is a running median filter, as we used in this paper.[3,20–27] Other possibilities are methods that fit more complex curves over the existing data. After applying the methods, a new dataset is created, and used in further analytical steps.

The last category consists of methods that replace the data signal with a function that is used directly in a statistical model. For example, spline functions are used instead of raw data in a model, or a joint model is constructed. This third type of method was not usable in this paper, because we required the processed blood pressure data to quantify hypotension in order to study the relation between hypotension and postoperative myocardial injury.

Despite our efforts to find artifact filtering methods in literature, the number of papers including methods is limited (38 papers) and the methods were quite heterogeneous. We looked for papers that mentioned artifacts in the title or abstract, which resulted in papers of which artifact filtering was an important element of the study (e.g., the development of an artifact filter). How often these

methods were used in practice cannot be extrapolated from this systematic search. It should be noted that in contrast with the title screening, the final identification and classification of the artifact filtering methods was performed by one author only. The selection of methods applicable to anesthesia information management system data was done by the research team, but was not predefined in the review protocol. Alternatively, we could have looked for the filters being applied in research practice, for example by searching for all papers studying intraoperative hypotension and outcomes, and then carefully studying the Methods sections. However, we expect that this would have resulted in a lower yield of artifact filtering methods. From personal experience as researchers in this field, we noticed that the methods for handling artifacts are typically not (extensively) reported.

Strengths and limitations

Our study is one of the first to explore the impact of artifacts on clinical research using anesthesia information management system data by applying different artifact filtering methods on real clinical data. We focused on different artifact filter methods and how they influence estimated associations. These methods were identified with a systematic search. Not only did we vary artifact filtering methods, we also studied different hypotension measures and varying hypotension definitions to get a complete picture. Our detailed analysis places the issue of artifacts into perspective, and the reader can base the methodologies for filtering artifacts in future work on these findings.

One limitation of our study is that we have chosen only one type of measurement (blood pressure) and one type of outcome. It is unclear to which extent we can generalize these findings to other physiologic data or research on other subjects. Second, we could not use every artifact filtering method found. More methods would have been applicable, if the anesthesia information management system data was more granular than data with 1-min intervals. However, the vital signs in our anesthesia information management system are based on raw data measured at 5-s intervals, of which the median per min is stored. This meant the data we used was already processed for artifacts by the anesthesia information management system software. Third, we could not compare the artifact filtering methods with a manual identification and exclusion of artifacts. For the purpose of our methods, comparison of a large dataset was required, but this hindered manual annotation of the data. Additionally, it would be questionable to do this in a retrospective

manner.[2,5] In the current study, we only included patients in whom invasive blood pressure monitoring was applied, resulting in a high proportion of high-risk surgery patients (table 1). In this subpopulation of the entire cohort, the incidence of postoperative myocardial injury was greater than previously reported.[9,11] nevertheless, this data were used as an illustration and were not used to estimate the true association between intraoperative and postoperative myocardial injury. Fourth, the effect of filtering artifacts could in theory be different within subgroups of procedures or patients. We did not adjust the association between intraoperative hypotension and postoperative myocardial injury for confounding, nor have we performed subgroup analyses. We considered this beyond the scope of the study. Consequently, as we have studied the effect of artifact filtering methods in a cohort with strict inclusion criteria, one cannot simply generalize these findings to a broader selection of procedures.

Fifth, we made other methodologic choices that could have influenced the association between hypotension and postoperative myocardial injury. For example, we chose linear interpolation between measurements,[12] instead of other methods such as last value carried forward. In a post hoc analysis we excluded cases with big gaps (greater than 15 min without data), which resulted in a tiny—although systematic—increase (about 0.02) in the odds ratio estimate. We therefore decided to not exclude any cases with gaps in blood pressure measurements, or cases in which a significant amount of measurements were removed by artifact filtering. These explorations suggested that the effect of bias (of gap removal) on the estimated odds ratio is minimal. In general, the variance of blood pressure measurements will be underestimated in linear interpolation, so if the primary interest is to estimate variability, then one should be more careful with interpolation.

Finally, this study did not cover all possible artifact filtering methods or combinations of methods. The aim of this study was to illustrate what happens when different filters are chosen, rather than to identify the best method. This best method is highly situational, depending on the type of data, the protocols used during anesthesia, the hardware used, and the way data is stored in an anesthesia information management system. Future (experimental) studies should aim to find reliable, generalizable methods for filtering artifacts in large anesthesia information management system databases.

Implications

It is hard to estimate the true association between hypotension and postoperative myocardial injury in our cohort, because we do not know which artifact filtering method is closest to the truth.[13] Even if we would have had a fully annotated dataset where artifacts are marked manually, one might still question whether this gives the correct estimate of the effect of intraoperative hypotension on postoperative myocardial injury.[28,29] Despite the fact that we cannot identify the best method for artifact handling, the use of different artifact methods in one paper, like we have done, is not generally advised. A researcher should choose one method, thereby explicitly defining the outcome of interest. Preferably, the outcome should be comparable to similar research, but this does not necessarily imply that the same artifact filtering should be used. The choice of artifact filtering method will depend on the nature of the data and the type of hypotension measure, as some measures are more sensitive to artifacts than others.

Over time, improvements in artifact handling of anesthesia monitoring systems may result in better research data that contains fewer artifacts. This will decrease the need for postprocessing data for research. Currently, monitor systems cannot prevent all types of artifacts in anesthesia information management system data. For example, an anesthesia monitor has no information available on the correct placement of sensors and the resulting blood pressure will not be recognized as an artifact by the anesthesia monitor artifact algorithm.

Overall conclusion

Although different artifact filtering methods yielded important differences in the quantification of intraoperative hypotension, we did not see a profound effect of these methods on the effect measures of the association between intraoperative hypotension and postoperative myocardial injury. It seems that the variation resulting from artifacts is smaller than the effect of the choice of hypotension measure or the chosen hypotension threshold. Nevertheless, the way one deals with artifacts may add to the reproducibility and comparability of intraoperative hypotension research. It seems wise to carefully consider how to handle artifacts in research using intraoperative physiologic data obtained from anesthesia information management systems. Authors should describe the chosen methodology for artifact filtering in detail.

Research Support

Support was provided solely from institutional and/or departmental sources.

Competing Interests

The authors declare no competing interests.

References

1. Martin M, Kheterpal S: Artéfacts de monitoring et recherche dans les grandes bases de données: Ce que vous ignorez pourrait vous nuire. Can J Anesth 2012; 59:823–6
2. Kool NP, Waes J a R van, Bijker JB, Peelen LM, Wolfswinkel L van, Graaff JC de, Klei W a van: Artifacts in research data obtained from an anesthesia information and management system. Can J Anaesth 2012; 59:833–41
3. Verduijn M, Peek N, Keizer NF de, Lieshout EJ van, Pont ACJM de, Schultz MJ, Jonge E de, Mol BAJM de: Individual and Joint Expert Judgments as Reference Standards in Artifact Detection. J Am Med Informatics Assoc 2008; 15:227–34
4. Cunningham S, Symon AG, McIntosh N: The practical management of artifact in computerised physiological data. Int J Clin Monit Comput 1994; 11:211–6
5. Hoorweg A-LJ, Pasma W, Wolfswinkel L van, Graaff JC de: Incidence of Artifacts and Deviating Values in Research Data Obtained from an Anesthesia Information Management System in Children. Anesthesiology 2017
6. Cao C, Kohane IS, McIntosh N: Artifact detection in cardiovascular time series monitoring data from preterm infants. Proc AMIA Symp 1999:207–11
7. Batista-Foguet JM, Coenders G, Arts Ferragud M: Using structural equation models to evaluate the magnitude of measurement error in blood pressure. Stat Med 2001; 20:2351–68
8. Maslove DM, Dubin JA, Shrivats A, Lee J: Errors, Omissions, and Outliers in Hourly Vital Signs Measurements in Intensive Care. Crit Care Med 2016; 44:e1021–30
9. Waes J a R van, Nathoe HM, Graaff JC de, Kemperman H, Borst GJ de, Peelen LM, Klei W a van: Myocardial injury after noncardiac surgery and its association with short-term mortality. Circulation 2013; 127:2264–71
10. Bijker JB, Klei WA van, Kappen TH, Wolfswinkel L van, Moons KGM, Kalkman CJ: Incidence of intraoperative hypotension as a function of the chosen definition: literature definitions applied to a retrospective cohort using automated data collection. Anesthesiology 2007; 107:213–20
11. Vernooij LM, Klei WA van, Machina M, Pasma W, Beattie WS, Peelen LM: Different methods of modelling intraoperative hypotension and their association with postoperative complications in patients undergoing non-cardiac surgery. Br J Anaesth 2018; 120:1080–9

12. Salmasi V, Maheshwari K, Yang D, Mascha EJ, Singh A, Sessler DI, Kurz A: Relationship between Intraoperative Hypotension, Defined by Either Reduction from Baseline or Absolute Thresholds, and Acute Kidney and Myocardial Injury after Noncardiac Surgery: A Retrospective Cohort Analysis. Anesthesiology 2017; 126:47–65
13. Wesselink EM, Kappen TH, Torn HM, Slooter AJC, Klei WA van: Intraoperative hypotension and the risk of postoperative adverse outcomes: a systematic review. Br J Anaesth 2018; 121:706–21
14. Sun LY, Chung AM, Farkouh ME, Diepen S van, Weinberger J, Bourke M, Ruel M: Defining an Intraoperative Hypotension Threshold in Association with Stroke in Cardiac Surgery. Anesthesiology 2018; 129:440–7
15. Tronstad C, Omenås IN, Rosseland LA: An improved artifact removal algorithm for continuous cardiac output and blood pressure recordings. Conf Proc . Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf 2015; 2015:5968–72
16. Cao H, Norris P, Ozdas A, Jenkins J, Morris JA: A Simple Non-physiological Artifact Filter for Invasive Arterial Blood Pressure Monitoring : a Study of 1852 Trauma ICU Patients 2006:0–3
17. Megjhani M, Terilli K, Martin A, Velazquez A, Claassen J, Roh D, Agarwal S, Smielewski P, Boehme AK, Schmidt JM, Park S: Deriving the PRx and CPPopt from 0 . 2-Hz Data : Establishing Generalizability to Bedmaster Users 2018; 126:0–3
18. Sottas CE, Cumin D, Anderson BJ: Blood pressure and heart rates in neonates and preschool children : an analysis from 10 years of electronic recording 2016; 26:1064–70
19. A.T. R, L. C, J. R: The association between vital signs and major hemorrhagic injury is significantly improved after controlling for sources of measurement variability. J Crit Care 2012; 27:533.e1-533.e10
20. Görges M, West NC, Karlsdóttir E, Ansermino JM, Cassidy M, Lauder GR: Developing an objective method for analyzing vital signs changes in neonates during general anesthesia. Paediatr Anaesth 2016; 26:1071–81
21. Wilson S, Bowyer A, Harrap SB: Characterizing the continuously acquired cardiovascular time series during hemodialysis, using median hybrid filter preprocessing noise reduction. Med Devices (Auckl) 2015; 8:65–70
22. Borowski M, Siebig S, Wrede C, Imhoff M: Reducing false alarms of intensive care online-monitoring systems: an evaluation of two signal extraction algorithms. Comput Math Methods Med 2011; 2011:143480

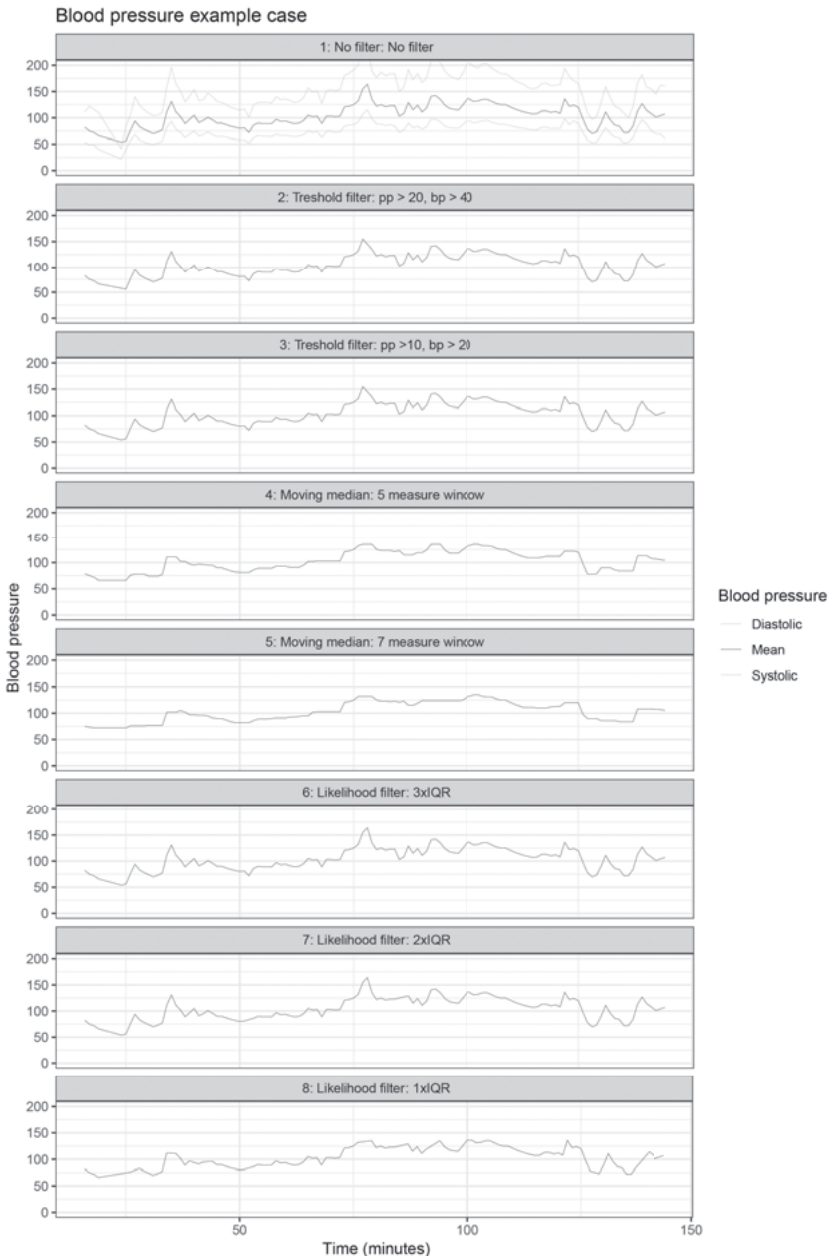
23. Yang P, Dumont GA, Ansermino JM: Sensor fusion using a hybrid median filter for artifact removal in intraoperative heart rate monitoring. J Clin Monit Comput 2009; 23:75–83
24. Clifford GD, Long WJ, Moody GB, Szolovits P: Robust parameter extraction for decision support using multimodal intensive care data. Philos Trans A Math Phys Eng Sci 2009; 367:411–29
25. Charbonnier S, Becq G, Biot L: On-line segmentation algorithm for continuously monitored data in intensive care units. IEEE Trans Biomed Eng 2004; 51:484–92
26. Hoare SW, Asbridge D, Beatty PCW: On-line novelty detection for artefact identification in automatic anaesthesia record keeping. Med Eng Phys 2002; 24:673–81
27. Jakob S, Korhonen I, Ruokonen E, Virtanen T, Kogan A, Takala J: Detection of artifacts in monitored trends in intensive care. Comput Methods Programs Biomed 2000; 63:203–9
28. Reich DL, Wood RK, Mattar R, Krol M, Adams DC, Hossain S, Bodian C a: Arterial blood pressure and heart rate discrepancies between handwritten and computerized anesthesia records. Anesth Analg 2000; 91:612–6
29. Wax DB, Beilin Y, Hossain S, Lin HM, Reich DL: Manual editing of automatically recorded data in an anesthesia information management system. Anesthesiology 2008; 109:811–5
30. Glynn RJ, L'Italien GJ, Sesso HD, Jackson EA, Buring JE: Development of predictive models for long-term cardiovascular risk associated with systolic and diastolic blood pressure. Hypertension 2002; 39:105–10
31. Tsien CL, Kohane IS, McIntosh N: Building ICU artifact detection models with more data in less time. Proceedings AMIA Symp 2001:706–10
32. Gostt RK, Rathbone GD, Tucker AP: Real-time pulse oximetry artifact annotation on computerized anaesthetic records. J Clin Monit Comput 2002; 17:249–57
33. Stegle O, Fallert S V, MacKay DJC, Brage S: Gaussian process robust regression for noisy heart rate data. IEEE Trans Biomed Eng 2008; 55:2143–51
34. Ansermino JM, Daniels JP, Hewgill RT, Lim J, Yang P, Brouse CJ, Dumont GA, Bowering JB: An evaluation of a novel software tool for detecting changes in physiological monitoring. Anesth Analg 2009; 108:873–80
35. Dosani M, Lim J, Yang P, Brouse C, Daniels J, Dumont G, Ansermino JM: Clinical evaluation of algorithms for context-sensitive physiological monitoring in children. Br J Anaesth 2009; 102:686–91

36. Caicedo A, Huffel S Van: Weighted LS-SVM for function estimation applied to artifact removal in bio-signal processing. Conf Proc . Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf 2010; 2010:988–91
37. Thomas L, Stefanski L, Davidian M: A moment-adjusted imputation method for measurement error models. Biometrics 2011; 67:1461–70
38. Stirnemann JJ, Comte F, Samson A: Density estimation of a biomedical variable subject to measurement error using an auxiliary set of replicate observations. Stat Med 2012; 31:4154–63
39. Graaff JC de, Bijker JB, Kappen TH, Wolfswinkel L van, Zuithoff NPA, Kalkman CJ: Incidence of intraoperative hypoxemia in children in relation to age. Anesth Analg 2013; 117:169–75
40. Ganesh SK, Chasman DI, Larson MG, Guo X, Verwoert G, Bis JC, Gu X, Smith A V, Yang M-L, Zhang Y, Ehret G, Rose LM, Hwang S-J, Papanicolau GJ, Sijbrands EJ, Rice K, Eiriksdottir G, Pihur V, Ridker PM, Vasan RS, Newton-Cheh C, Global Blood Pressure Genetics Consortium, Raffel LJ, Amin N, Rotter JI, Liu K, Launer LJ, Xu M, Caulfield M, Morrison AC, Johnson AD, Vaidya D, Dehghan A, Li G, Bouchard C, Harris TB, Zhang H, Boerwinkle E, Siscovick DS, Gao W, Uitterlinden AG, Rivadeneira F, Hofman A, Willer CJ, Franco OH, Huo Y, Witteman JC, Munroe, PB, Gudnason V, Palmas W, van Duijn C, Fornage M, Levy D, Psaty BM, Chakravarti A: Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations. Am J Hum Genet 2014; 95:49–65
41. Terrill PI, Dakin C, Hughes I, Yuill M, Parsley C: Nocturnal oxygen saturation profiles of healthy term infants. Arch Dis Child 2015; 100:18–23
42. Cross TJ, Keller-Ross M, Issa A, Wentz R, Taylor B, Johnson B: The Impact of Averaging Window Length on the"Desaturation Indexes during Overnight Pulse Oximetry at High-Altitude". Sleep 2015; 38:1331–4
43. Mort AJ, Fitzpatrick D, Wilson PMJ, Mellish C, Schneider A: Lightweight physiologic sensor performance during pre-hospital care delivered by ambulance clinicians. J Clin Monit Comput 2016; 30:23–32
44. Hwang Y-T, Wang C-C, Wang CH, Tseng Y-K, Chang Y-J: Joint model of multiple longitudinal measures and a binary outcome: An application to predict orthostatic hypertension for subacute stroke patients. Biom J 2015; 57:661–75
45. Zahari M, Lee DS, Darlow BA: Algorithms that eliminate the effects of calibration artefact and trial-imposed offsets of Masimo oximeter in BOOST-NZ trial. J Clin Monit Comput 2016; 30:669–78

46. Fathabadi OS, Gale T, Wheeler K, Plottier G, Owen LS, Olivier JC, Dargaville PA: Hypoxic events and concomitant factors in preterm infants on non-invasive ventilation. J Clin Monit Comput 2017; 31:427–33
47. Chen L, Dubrawski A, Wang D, Fiterau M, Guillame-Bert M, Bose E, Kaynar AM, Wallace DJ, Guttendorf J, Clermont G, Pinsky MR, Hravnak M: Using Supervised Machine Learning to Classify Real Alerts and Artifact in Online Multisignal Vital Sign Monitoring Data. Crit Care Med 2016; 44:e456-63
48. Georgatzis K, Lal P, Hawthorne C, Shaw M, Piper I, Tarbert C, Donald R, Williams CKI: Artefact in Physiological Data Collected from Patients with Brain Injury: Quantifying the Problem and Providing a Solution Using a Factorial Switching Linear Dynamical Systems Approach. Acta Neurochir Suppl 2016; 122:301–5
49. Nixon GM, Davey MJ, Weichard AJ, Horne RSC: Oximetry for suspected obstructive sleep apnea-Does removal of awake data affect the result? Pediatr Pulmonol 2016; 51:1409–13
50. Hravnak M, Chen L, Dubrawski A, Bose E, Clermont G, Pinsky MR: Real alerts and artifact classification in archived multi-signal vital sign monitoring data: implications for mining big data. J Clin Monit Comput 2016; 30:875–88
51. Kim H, Lee S-B, Son Y, Czosnyka M, Kim D-J: Hemodynamic Instability and Cardiovascular Events After Traumatic Brain Injury Predict Outcome After Artifact Removal With Deep Belief Network Analysis. J Neurosurg Anesthesiol 2018; 30:347–53
52. Colopy GW, Tingting Zhu, Clifton L, Roberts SJ, Clifton DA: Likelihood-based artefact detection in continuously-acquired patient vital signs. Conf Proc. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf 2017; 2017:2146–9
53. Wellington G, Elder D, Campbell A: 24-Hour Oxygen Saturation Recordings in Preterm Infants: Editing Artefact. Acta Paediatr Int J Paediatr 2018; 107:1362–9

Supplemental Material

This example illustrates the effect of different artifact filtering methods on a blood pressure signal of one of the procedures in our cohort. Eight different methods are illustrated (including no filter). These methods are also described in the paper.



S1: examples of cases with different filters applied. This example illustrates the effect of different artifact filtering methods on a blood pressure signal of one of the procedures in our cohort. Eight different methods are illustrated (including no filter). These methods are also described in the paper.

6

Artifacts annotations in anesthesia blood pressure data by man and machine

Wietze Pasma
Esther M. Wesselink
Stef van Buuren
Jurgen C. de Graaff
Wilton A. van Klei

Accepted for publication in Journal of Clinical Computing and Monitoring 2020

Abstract

Purpose

Physiologic data from anesthesia monitors are automatically captured. Yet erroneous data is stored in the process as well. While this is not interfering with clinical care, research can be affected. Researchers should find ways to remove artifacts. The aim of the present study was to compare different artifact annotation strategies, and to assess if a machine learning algorithm is able to accept or reject individual data points.

Methods

Non-cardiac procedures requiring invasive blood pressure monitoring were eligible. Two trained research assistants observed procedures live for artifacts. The same procedures were also retrospectively annotated for artifacts by a different person. We compared the different ways of artifact identifications and modelled artifacts with three different learning algorithms (lasso restrictive logistic regression, neural network and support vector machine).

Results

In 88 surgical procedures including 5711 blood pressure data points, the live observed incidence of artifacts was 2.1 % and the retrospective incidence was 2.2%. Comparing retrospective with live annotation revealed a sensitivity of 0.32 and specificity of 0.98. The performance of the learning algorithms which we applied ranged from poor (kappa 0.053) to moderate (kappa 0.651).

Conclusion

Manual identification of artifacts yielded different incidences in different situations, which were not comparable. Artifact detection in physiologic data collected during anesthesia could be automated, but the performance of the learning algorithms in the present study remained moderate. Future research should focus on optimization and finding ways to apply them with minimal manual work. The present study underlines the importance of an explicit definition for artifacts in database research.

Introduction

Physiological data captured by anesthesia monitors are used for medical record keeping during anesthesia. Such data are stored in Anesthesia Information Management System (AIMS) databases along with other anesthesia record keeping data. These rich databases are widely used for clinical research, since the data are obtained without much extra effort or altering the clinical workflow. However, automatically collected monitor data often include erroneous data that are not reviewed before they are stored, which may bias research results.[1,2] In daily practice the anesthesiologist ignores artifacts based on other information available and therefore clinical anesthesia care is not affected. For example, artifacts in the ECG signal caused by detached electrodes ('asystole') can be ignored in the operating room based on other monitoring (normal invasive blood pressure signal). When these same data are used for research purposes, the context of the procedure is lost, and it is harder to distinguish which measurements are true and which are artifacts.[3,4]

Usually researchers come up with a definition of artifacts, and apply this definition to the data to correct errors. For example, values above and below a certain threshold are defined as artifacts and consequently removed from the data. Several other solutions for artifact removal are available[5] and effectiveness and accuracy of these methods will depend on the situation in which the data was collected. There is no consensus on which method for artifact removal should be used in AIMS database research. Thus far the removal of artefactual data live by hand (with presence in the operating room) seems to be the gold standard, to which other methods are compared.[6–9] Manual cleaning of artefactual data is not only a cumbersome and time consuming process, but could also depend on the situation in which the data was collected, similarly to automatic filtering methods. Factors that can affect artifact annotation are for example: the time when the data was annotated (during the anesthesia procedure or afterwards) the location (presence in the operating room or remote location) and who annotated the data (i.e. an anesthesiologist, or a researcher). The pre-existing definition of artifacts in the rater's mind or defined in a study protocol will influence which values will be marked as artifacts. Therefore manual annotations and models based on these manual annotations are difficult to generalize.[7,10]

The aim of the present study was to compare different artifact annotation strategies, and subsequently to assess if a learning algorithm would be able to simulate the decision to accept or reject individual vital sign data points. We used invasive blood pressure measurements as an example and hypothesized that regardless of different annotation methods, an algorithm could be trained to perform with comparable accuracy. Such algorithm learning strategy could standardize automated artifact removal to clean data for clinical research.

Methods

Live observation

Two research assistants (last year medical students) were trained by an anesthesiologist, to observe and annotate anesthesia procedures, during a time period of 11 weeks (between July 29th and October 11th) in the University Medical Center Utrecht. The local ethics committees approved the protocol and waived the need for informed consent (University Medical Center Utrecht Medical Research Ethics Committee, protocol no. 19-629). At the start of each workday, the assistant identified non-cardiac procedures in adults with planned invasive blood pressure monitoring. Each measurement session covered a part of the anesthesia procedure of at least one hour, to ensure that a mix of procedures was sampled. Procedures were preferably visited after induction or before end of surgery. The visiting order was not randomly determined, rather procedures were visited in sequence (i.e. when an observation was finished, the research assistant would identify the next eligible procedure to visit). The research assistant registered observations on a device (laptop) that was not connected to the operating room equipment. We used Behavioral Observation Research Interactive Software (BORIS, http://www.boris.unito.it, Torino, Italy) software to record live observations during anesthesia. This software package allows for swift registration of observations, using a keystroke per type of observation, ensuring fine granularity in the data.[11]

Invasive blood pressure was measured with an IntelliVue monitoring system (type MP70, X2 multimeasurement module; Philips, Germany). To mark the beginning of a registration, the arterial catheter was flushed, which generated a flush artefact in the waveform recordings, which was used later. Observers were instructed to document the start and end of any disturbance of the waveform signal, displayed on the patient monitor. At the same time, the reason for this artifact period was documented. When the observer was uncertain about the artifact events, he was

permitted to discuss the artifact with the clinician responsible for the anesthetic procedure. The different artifact reasons for blood pressure were flush, blood sampling, sensor issues or movement of the patient, simultaneous non-invasive blood pressure measurement and height of the pressure sensor. The observer described this reason as a free text comment. When the observer was not able to categorize the reason of an artifact, this was later discussed with the research team and categorized.

The BORIS live registration software was not able not register the true observation time, based on the true time of the AIMS database (Anstat, version 2.1, 2019, Carepoint, Ede, The Netherlands). Therefore we looked up the flush event in the stored waveform data. With the registered flush time at the beginning of the observation and the actual flush time we synchronized the live observations with the data points in the AIMS database. Waveform data was analyzed using SignalBase, version 10.0.0 (legal copyright: UMC Utrecht MTKF, Utrecht, the Netherlands). SignalBase was developed to review and analyze raw waveform data as stored in the AIMS database.

Data registration and retrospective annotation

In the institutional AIMS one data point for each measurement minute of invasive blood pressure is stored. This value is calculated by taking the median of the previous measurement minute (which encompasses 12 data points outputted by the patient monitor at 1/5 Hz). The resulting data points (1/60 Hz) were annotated retrospectively by one of the researchers (E.W.). This annotation was done using an interactive *R shiny* application designed for this study (https://github.com/wietze314/annotate-vital-signs), which displayed data points from the AIMS database. The invasive blood pressure measurements were annotated, with the complete health record available, as an additional reference, but blinded by the artifact identification done by the live observant. Therefore the observer was also aware of events and medication administrations during anesthesia. Besides artifact annotations, the application also collected meta-data such as the start and end time of the annotation process per observation.

Data preparation and definition of artifacts

In order to compare the live observations to the retrospective annotations, we linked both datasets to the vital signs data in the AIMS database. During live

observation the time period, in which the monitor displayed artifactual data, was registered. The data points stored in the AIMS database were based on a period of one minute of measurements. To combine these data, we calculated how much overlap each of the measurement minutes had with artifact periods. If there was any overlap (i.e. the AIMS data point was based on waveform data with artifacts), the data point was defined as an artifact (definition 1). Additionally, to generate a more specific artifact definition, we defined a data point as an artifact when there was more than 30 seconds of artifact during a measurement minute (definition 2). Furthermore, we marked individual data points as artifacts retrospectively, based on the available AIMS data (definition 3). The concept of data preparation is illustrated in Figure 1.

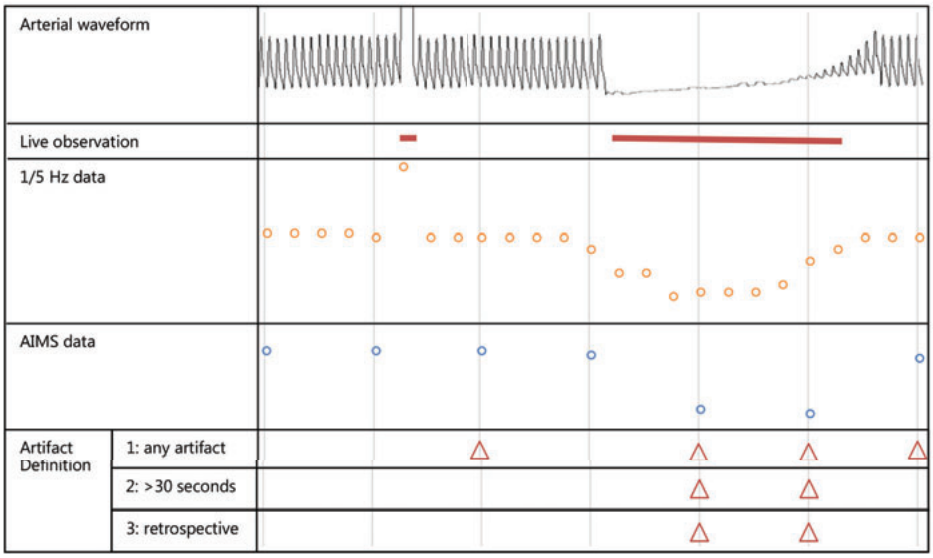


Figure 1: A simplified display of monitor and anesthesia information management system (AIMS) database data. Live artifact observations (red line) were based on the arterial waveform displayed on the patient monitor. Each minute (vertical lines) the AIMS software stored a data point (blue), which was based on the previous measurement minute (1/5 Hz data) (orange data points). The live observations were translated to artifacts by two definitions. Finally AIMS data points were retrospectively annotated.
Definition 1: When there was any artifact during the measurement minute, the data point was identified as an artifact.
Definition 2: When there was more than 30 seconds of artifact, during the measurement minute, the data point was identified as an artifact.
Definition 3: Retrospectively identified artifacts, according to stored AIMS data points.
Definition 4: The 1/5 Hz data points (orange) were considered artifacts, when they fell within a artifact period (red lines)

Apart from the data points stored each minute (1/60 Hz), we also collected 1/5 Hz data from a subsample of the cases, of which these higher resolution data were

automatically stored in the AIMS database. We could not collect these data from all cases, because it required a change in settings of the AIMS software, which was set in only a part of the operating rooms for the purpose of this study. A 1/5 Hz data point was defined as an artifact, when this data point was within a live observed artifact period (definition 4) (Figure 1).

In addition to the data collection of artifact periods, the observer also noted the cause of these artifactual periods. In the rare event of a data point which was influenced by multiple artifacts, the cause of artifact for that data point was determined by a majority vote, i.e. the cause that represented the most seconds of artifact was chosen. For example if within a measurement minute there was an artifact of 10 seconds because of blood sampling and subsequently an artifact of 30 seconds caused by manipulation of the blood pressure sensor, the data point was classified under the latter cause (Sensor issues or movement).

Blood pressure processing

Our aim was to predict artifacts in AIMS data, using data normally available for retrospective database research. Therefore, we chose to only use vital signs data points (blood pressure and heart rate) to extract features for algorithm training. The feature types that we extracted and calculated were: systolic blood pressure, diastolic blood pressure, mean blood pressure, heart rate, pulse pressure (systolic – diastolic blood pressure), ratios between heartrate and blood pressure (systolic blood pressure divided by heartrate and diastolic blood pressure divided by heartrate), ratio between systolic and mean arterial blood pressure, and ratio between mean and diastolic arterial blood pressure. For each of the aforementioned features, the differences between the current data point and the five previous and five next data points were calculated. Thus in total eleven features were generated per feature type. Additionally, the median and the mean was calculated from the current data point and the five preceding and five following data points (11 data points), for heartrate, diastolic, systolic and mean blood pressure. Also the difference and the relative difference (difference divided by the median or mean) between the current data point and the median and mean was calculated. This process resulted in 111 features to present to a learning algorithm.

The same procedure was used for 1/5 Hz vital signs collected from the AIMS database, in a subset of the cohort. Only now the 15 preceding and 15 following

data points for each data points were used, with an interval of 20 seconds between data points (i.e. skipping 3 measurements each time). This in turn, generated 291 features in total.

Statistical analysis

Incidences of artifacts in different artifact annotation definitions were calculated, as well as differences between both observers. We then compared live and retrospective observations by creating contingency tables and calculating the sensitivity, specificity and positive predictive value. In this, the live artifact annotations (definition 1 and 2) were used as the reference. Finally we calculated the time which was used to annotate the data for the different artifact definition.

Learning algorithms

For training learning algorithms, we considered each AIMS data point as an independent observation. The features were calculated using the surrounding data of each data point, thus providing the algorithms also with information of changes in time. All data points were first randomly assigned to the training and test set, with probability of 0.8 and 0.2 respectively. We generated a different training and test set for each learning algorithm method and each artifact definition, using different random seeds.

We used three different (machine) learning algorithms to model artifacts in invasive blood pressure data, i.e. lasso penalized logistic regression, a single layer neural network and a support vector machine.[12] First, we optimized the chosen learning algorithm. The training set was used to train the algorithm, and to tune the hyper parameters, with a 4-fold cross-validation. The optimal performing set of hyper parameters was chosen based on the Kappa statistic. We chose Kappa as a performance measure over accuracy, because the incidence of the outcome (artifact) was rare. Kappa corrects for agreement by chance, and is more informative for highly skewed data.[13] We trained and optimized all four artifact definitions separately, thus each optimal model had dedicated hyper parameters. All algorithms were trained using the caret package in R.[14] We expressed performance of the different algorithms in sensitivity and specificity and positive predictive value, based on the reference artifact data, which we manually collected. Finally, we evaluated performance of the learning algorithm on the test set, which was kept separated from the algorithm learning procedure and cross-validation.

The *glmnet* method was used to train a lasso restrictive logistic regression.[15] The hyper parameter lambda was optimized, which defines how much the estimates are penalized, and therefor determines the generalizability of the model. Alpha was kept constant at one, which means that lasso regression is performed. A neural network was trained using the *nnet* method.[16] The neural network consisted of one hidden layer, which size we optimized with cross-validation, testing 2 to 20 units. In addition the hyper parameter weight decay was optimized setting its value from 10^{-7} to 10. The weight decay parameter determines how much estimates are penalized, and therefor determine how generalizable a model will be. Finally, we trained a support vector machine with a radial basis function or Gaussian kernel,[17] using the *svmRadial* (*e1071* R package) method. The hyper parameter C and sigma were optimized with cross-validation. The C parameter defines how much estimates are penalized, and therefore determines the generalizability of the model, where a low C means more generalizable. C was varied from $5*10^{-4}$ to 10^{3}. The sigma parameter (or gamma parameter) determines the reach of each data point, which influences which observations determine the decision boundary of the support vector machine. Low sigma will result in a more linear decision boundary and a lower variance model than when a higher sigma is used. We varied sigma from $5*10^{-4}$ to 0.2.

As a post-hoc analysis, we collected extra procedures which we annotated retrospectively, to see if the performance of learning algorithms might improve. These cases were randomly selected from January 1st to June 1st 2019. We only considered non-cardiac and non-thoracic surgery in adults. When there were issues with documentation (i.e. the health record was incomplete) the procedure was also excluded. In each selected procedure, a period of 60 minutes was randomly chosen for annotation. The middle of the procedure was preferred (higher probability of sampling), which was similar to the sampling strategy we used in the prospective cohort.

De-identified data collection and statistical analysis was performed with R (R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org, R version 3.5.1 (2018-07-02)).

Results

Cohort

In total we included 88 procedures, which summed up to a total of 5711 blood pressure data points. Additionally, 29476 blood pressure data points from 40 procedures captured at a higher frequency (1/5 Hz) were included (figure 2). Baseline characteristics for the observations are listed in table 1.

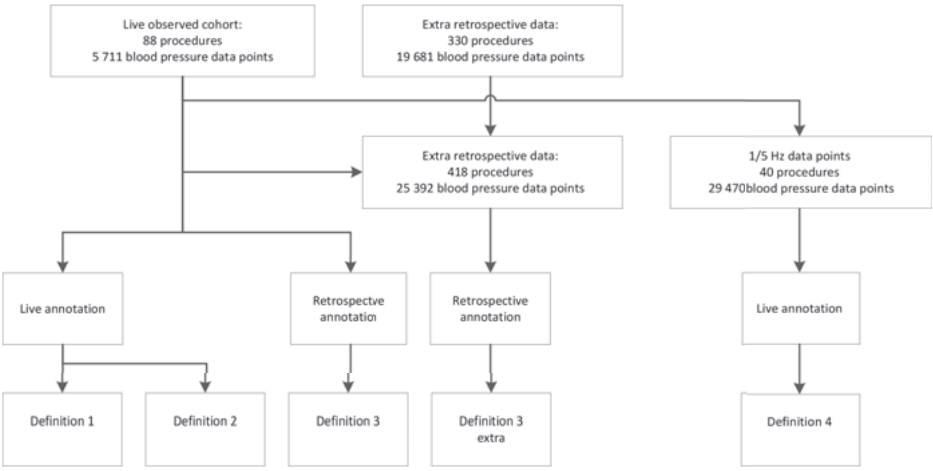


Figure 2: Flow diagram inclusion of observed procedures and which artifact definitions were applied. 88 procedures were live observed. 330 extra procedures were retrospectively collected for a post-hoc analysis.

Table 1: Baseline characteristics of observed procedures

Parameter	**N (%) or Median (IQR)**		
	All	Observer 1	Observer 2
Number of procedures	88	37 (100 %)	52 (100 %)
Age (yrs)	66 (56-74)	69 (57-76)	66 (55.5-74)
Durations of surgery (minutes)	344 (241-463)	327 (239-471)	356 (249-461)
Male	39 (44 %)	19 (51 %)	20 (38 %)
Weight (kg)	76 (68.6-86.7)	76 (69.3-86.7)	75 (67.4-85.5)
Surgical specialty			
ENT & Maxillofacial	15 (17 %)	8 (22 %)	7 (13 %)
General	30 (34 %)	11 (31 %)	19 (37 %)
Neurosurgery	23 (26 %)	10 (28 %)	13 (25 %)
Gynaecology & Urology	17 (19 %)	5 (14 %)	12 (23 %)
Other	3 (3 %)	2 (6 %)	1 (2 %)
ASA classification			
1	5 (6 %)	1 (3 %)	4 (8 %)
2	47 (53 %)	17 (47 %)	30 (58 %)
3 or 4	36 (41 %)	18 (50 %)	18 (35 %)

Reported values are number of procedures (percentage) or median and interquartile range (IQR). ASA American Society of Anesthesiologists. ENT Ear nose and throat

Artifacts

In 5711 blood pressure data points, 349 (6.1%) were based on data with an artifact, annotated by the observer (definition 1). According to definition 2 (at least 30 seconds of artifact), only 118 (2.1%) data points were identified as artifacts. Artifact incidence was 2.4% and 1.8% for observer 1 and observer 2, respectively. Retrospective artifact annotation (definition 3) yielded 124 (2.2%) artifacts (Table 2).

Table 2: Artifact incidence according to four artifact definitions

	N	(%)
Cohort blood pressure data points	5711	
Definition 1: Any artifact	349	(6.1)
Definition 2: >30 seconds artifact	118	(2.1)
Definition 3: retrospective annotation	124	(2.2)
1/5 Hz blood pressure data points	29470	
Definition 4: within artifact	761	(2.6)
Post-hoc additional data points	19681	
Definition 3: retrospective annotation	226	(1.1)

Within the 40 cases that contributed 1/5 Hz data, we identified 761 (2.6%) of 29476 data points as artifacts. For the post-hoc analysis we collected an additional 330 retrospective observations, with 19681 blood pressure data points, of which 226 (1.1%) were identified as artifacts. The median time spend for retrospective annotation of 418 observations was 25 seconds (IQR 13 - 47 seconds)).

Taking live annotation (definition 1) as the reference, retrospective annotation had a sensitivity of 0.14 and a specificity of 0.99, with a positive predictive value of 0.40. Comparing the more specific artifact definition 2 (at least 30 seconds of artifact) yielded a sensitivity of 0.32 and specificity 0.98, with a positive predictive value of 0.31. Table 3 lists both contingency tables and test parameters.

The most frequently reported cause of an artifact was sensor issues and/or movement of the patient (Table 4).

Table 3: Comparison of artifact annotations

	Definition 1: any artifact		**Definition 2: >30 seconds of artifact**	
Definition 3: Retrospective	**Artifact**	**No artifact**	**Artifact**	**No artifact**
Artifact	49	75	38	86
No artifact	298	5289	80	5507
Sensitivity	0.14		0.32	
Specificity	0.99		0.98	
Positive predictive value	0.40		0.31	

Contingency tables for live observed artifacts (definition 1 and 2), compared to retrospective annotation (definition 3)

Table 5: Learning algorithms, to predict artifacts in AIMS vital signs data

			4-fold cross-validation				Performance test dataset			
Artifact definition	**Function**	**Hyper parameters**	**Kappa**	**Sens**	**Spec**	**PPV**	**Kappa**	**Sens**	**Spec**	**PPV**
1: Any artifact	glmnet	alpha = 1, lambda = 2.78e-06	0.168	0.124	0.989	0.415	0.053	0.041	0.991	0.250
	nnet	size = 20, decay = 1e-07	0.166	0.240	0.939	0.209	0.087	0.158	0.941	0.123
	svmRadial	sigma = 0.001, C = 1000	0.216	0.204	0.974	0.329	0.201	0.200	0.969	0.304
2 : >30 seconds of artifact	glmnet	alpha = 1, lambda = 3.59e-05	0.285	0.241	0.992	0.401	0.066	0.062	0.992	0.100
	nnet	size = 8, decay = 0.01	0.226	0.277	0.979	0.221	0.118	0.154	0.976	0.129
	svmRadial	sigma = 5e-04, C = 1000	0.215	0.182	0.991	0.309	0.183	0.176	0.991	0.214
3 : Retrospective annotation	glmnet	alpha = 1, lambda = 2.78e-06	0.389	0.339	0.995	0.671	0.447	0.407	0.991	0.524
	nnet	size = 20, decay = 0.001	0.426	0.353	0.994	0.597	0.588	0.500	0.996	0.733
	svmRadial	sigma = 5e-04, C = 100	0.530	0.438	0.996	0.716	0.524	0.429	0.995	0.706
3: Retrospective (additional data points)	glmnet	alpha = 1, lambda = 4.64e-04	0.462	0.315	1.000	0.923	0.399	0.275	0.999	0.759
	nnet	size = 10, decay = 1	0.560	0.469	0.997	0.721	0.481	0.386	0.997	0.659
	svmRadial	sigma = 5e-04, C = 50	0.552	0.431	0.998	0.790	0.651	0.521	0.999	0.884
4: 1/5 Hz data	glmnet	alpha = 1, lambda = 2.78e-06	0.245	0.169	0.997	0.592	0.100	0.056	0.999	0.615
	nnet	size = 9, decay = 1	0.468	0.395	0.993	0.611	0.538	0.486	0.993	0.627
	svmRadial	sigma = 0.005, C = 50	0.616	0.497	0.998	0.840	0.631	0.518	0.997	0.830

glmnet: lasso regularized logistic regression, nnet: Neural network with one hidden layer, svmRadial: Support vector machine with Gaussian kernel, Sens: Sensitivity, Spec: Specificity, PPV: Positive predictive value. In this table the performance of the algorithm in the cross-validation and in the test dataset are presented.

Table 4: Artifact causes

	Definition			
	Definition 1: Any artifact		Definition 2: >30 seconds of artifact	
Cause of artifact	**Artifacts**	**Retrospective**	**Artifacts**	**Retrospective**
Blood sampling	40(12%)	8(20%)	23(19%)	5(22%)
Sensor issues or movement	136(39%)	3(2%)	23(19%)	3(13%)
Flush	9(3%)	1(11%)	1(1%)	1(100%)
Height of pressure sensor	56(16%)	23(41%)	37(31%)	20(54%)
Simultaneous NIBP	68(19%)	4(6%)	18(15%)	3(17%)
Other/Not specified	38(11%)	10(26%)	16(13%)	6(38%)

Artifact causes and retrospective identification. Results are presented as N (%). Each second column's percentage is calculated by dividing the retrospectively detected (true positives) by the total number of artifacts in that category. NIBP: non-invasive blood pressure.

Learning algorithms

For each of the four artifact definitions, three different machine learning algorithms were fitted. Performance ranged from poor (definition 1, lasso regularized logistic regression, kappa 0.053) to moderate (definition 3: retrospective annotation, neural network, kappa 0.588). Few normal data points were marked as artifacts (false positives), therefore the learning algorithms overall had a high specificity. When the amount of data presented to the learning algorithm, increased (definition 3 with additional data) the performance also increased. For example, for support vector machine, kappa increased from 0.524 to 0.651 (Table 5).

Discussion

Main findings

We compared different artifact annotation strategies of captured blood pressure data points in an AIMS database. Live annotated artifacts were frequently not identified as artifacts retrospectively (sensitivity of 0.32). The learning algorithms we subsequently developed to artificially identify artifacts were not able to adequately model artifacts which were annotated during live observations. Although the performance of these algorithms increased when retrospective annotations were modelled, the overall performance remained moderate.

Artifacts in invasive blood pressure measurements have different causes, such as movement or measurement technique artifacts. Some of the artifacts were short lasting and harder to pick up retrospectively, while others were longer lasting (Table 4). For example, movement artifacts according to artifact definition 1 were in only 2% of the cases identified as an artifact retrospectively (definition 3). On the other hand, the artifacts according to definition 2 (>30 seconds of artefactual signal) were retrospectively identified in 13% of the cases. In the present study, the AIMS used a calculated median of one minute of data to store data points. It therefore makes sense that short lasting artifacts have not resulted in artifactual data points within the AIMS database, which could be identified retrospectively. Nonetheless, we would have expected a larger difference in the effect described above. In addition, we found variation over different causes of artifacts in retrospective positively identified artifacts. These differences likely were a result of differences in information availability per situation. For example, from the AIMS record data points with systematic errors in blood pressure measurement due to the height of artery sensor placement were easier identified, than artifacts caused by movement.

In the present study we present different methods to manually define artifacts in AIMS data, and compare these different definitions with each other. Others have analyzed differences between artifact annotations, but comparisons were done to compare different raters, who received the same annotation task, i.e. retrospective annotation. The present study shows that it is not only important to describe who annotated data, but also when and how data points were marked as artifacts, in order to make research reproducible.[7]

We have prospectively collected data during a period of twelve weeks. This resulted in a reasonable quantity of observations. Nevertheless, the incidence of artifacts in the present study was quite low (2%). The amount of data available for the learning algorithms might thus have been too small. We observed procedures mainly in the maintenance phase of surgery, as we expected that it would be more complex to label artifacts precisely in the induction and emergence period where a lot of things happen at the same time. The artifact incidence was similar to what was previously found during maintenance in pediatric surgery, which was also lower than during induction or emergence.[3] Furthermore the type of surgery could have affected the incidence of artifacts, for example the cohort had a high portion of neurosurgery procedures, during which movement of the patient is limited and the surgical field is further away from the blood pressure sensor than other types of surgery.

In the present study, only one researcher annotated the data retrospectively, which can be considered a limitation. We could have improved quality of annotation when more than one researcher had annotated the data. On the other hand, because differences between these raters also need to be evaluated, the time invested in an extra person who annotated the data, would have been considerably more than twice the time which we spent thus far. Also the goal of this research was not to compare raters with each other, as has already been done previously.[7]

We used two definitions to translate the live observations to an artifact definition (definition 1 and 2), using the duration of the artifacts. Another approach could have been to combine the severity of the artifact, for example the deviance from baseline, with the duration of the artifact. In theory short extreme artifacts (e.g. flush events) can affect analysis differently than long but less extreme artifacts (e.g. height of the pressure sensor). In our situation the duration of artifacts was more important since our system stores the median of 12 consecutive blood pressure measurements. Therefore we only used the duration of artifacts, but in other situations this definition might be too limited.

We used two research assistants for live observations, which could result in differences in the way data was annotated. We observed minor differences in artifact incidence in each group of procedures, which were probably due to differences in procedure types (Table 1). The number of artifacts according to

retrospective annotation (definition 3) varied in a similar way, between these two subgroups (data not presented). Unfortunately, we have not performed a double-code observation to compare both observers adequately.

We have purposefully used only automatically collected physiologic data captured during the anesthetic procedure as source of features. We made this choice to ensure that resulting methodology and workflow will be generalizable, even when no other data than vital signs are available. This approach makes the methodology broadly applicable. On the other hand we tried to model a (human) decision, i.e. manual artifact identification, with limited information, from which the performance of the learning algorithms would have suffered. We saw that none of the learning algorithms performed well enough to apply for future research, as presented here. We showed in a post-hoc analysis that the performance could improve by adding additional data points. Nevertheless the information that was available to these algorithms was probably still too limited. To understand this concept better, future research could focus on adding not only more observations but also more features to the model, which are commonly available in databases used for research. For example patient characteristics, procedure type and medication administration or other events around the data point of interest could be added.

Implications

Before we can say anything about the implications of this study we first need to consider the definition of an artifact. Is every measurement in an AIMS database, based on a disturbed or a not perfect signal an artifact? Or does a live observed disturbance in a signal only produce an artifact, when the stored data point is different than what we expect for a patient at that particular time during anesthesia? But in the latter case, how do we define an expected value? These questions show that artifact annotation is a subjective matter, and a question of definition. It is important that researchers report what they considered to be artifacts, even when this process was done manually.

Despite this issue in defining artifacts, artifact annotation could still be automated using learning algorithms. The present study showed that this is not straightforward and might still require an investment of time to collect manually annotated data points for training. We live observed around 95 hours of anesthesia, while using retrospective data from 328 hours only improved the performance of the learning

algorithms marginally. Observing this much data live would have been very labor intensive and likely not feasible. In contrast, retrospectively annotating these data took us around three hours with a custom made registration application by a single person. This makes retrospective annotation better suitable to remove artifacts from research data, than live annotation.

Even though machine learning algorithms performed poorly in the present study, our approach is still insightful for those who want to apply similar annotation tools and models on their own AIMS data. Future research could focus on improving the performance and develop application of the methods presented in the present paper. To minimize time spend on manual data collection, we suggest optimizing this process using an active learning strategy. With this strategy, only data points are annotated, which contribute significantly to the learning algorithm. This could reduce the time spend on annotating data significantly.[18,19]

Conclusion

Identification of artifacts in invasive blood pressure measurements depends on the moment of annotations (live versus retrospective) and the person who annotated the data points. Nevertheless, these different artifact definitions could be modelled with learning algorithms in a similar way. The performance of these algorithms was poor in the present study and should be improved before applying in the future. A substantial amount of manually annotated data is still required to train these algorithms. As a positive by product of such an effort, researchers are forced to define explicitly what artifacts in their data are.

Conflict of interest

The authors declare that they have no conflict of interest.

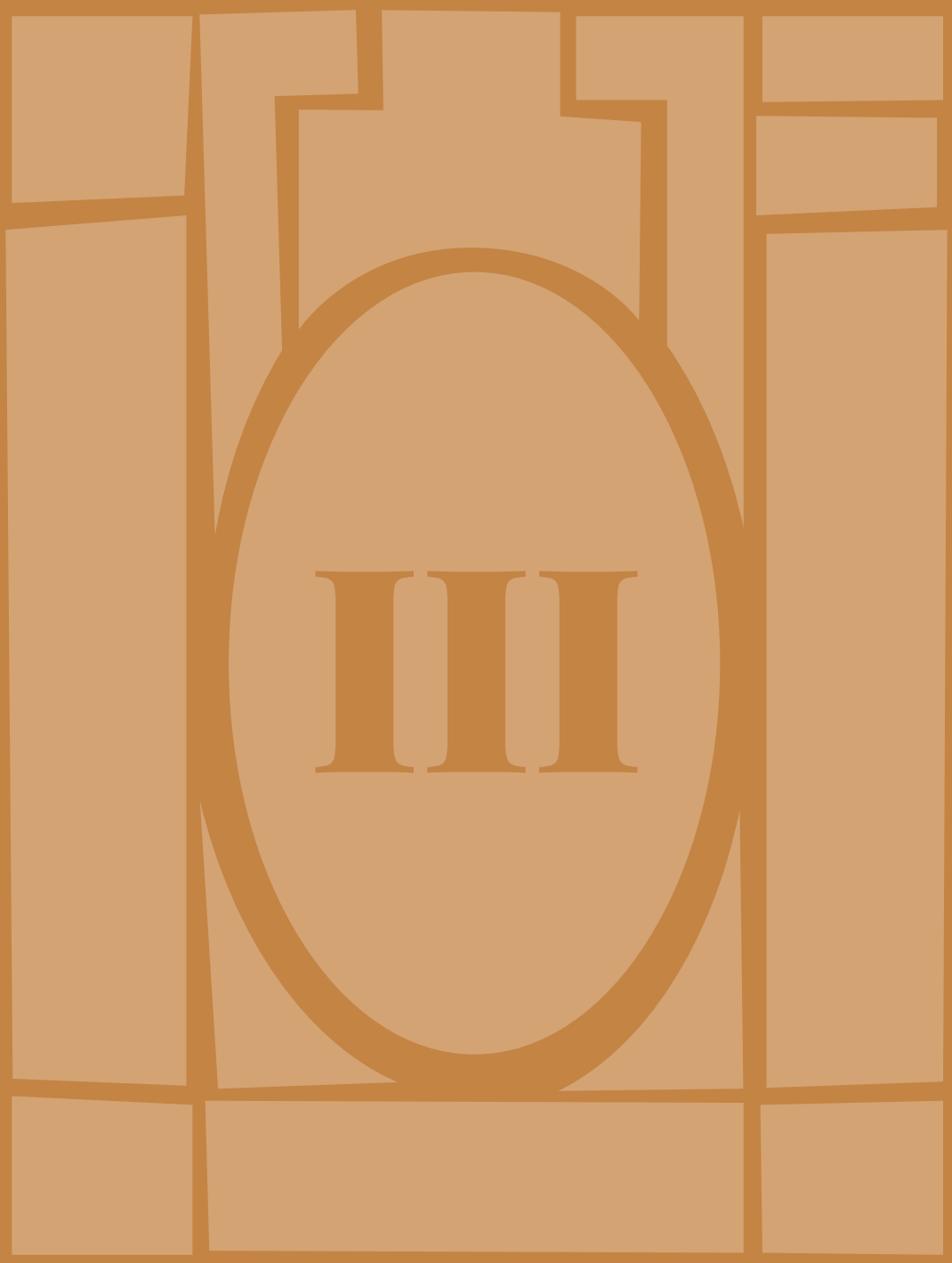
Acknowledgements

We like to acknowledge research assistants Wisse Beuken and Huib van Dijk for patiently observing the anesthesia procedures, and precisely collecting the data which were used in this publication.

References

1. Meyfroidt G, Güiza F, Ramon J, Bruynooghe M. Machine learning techniques to examine large patient databases. Best Pract Res Clin Anaesthesiol. 2009;23:127–43.
2. De Graaff JC, Engelhardt T. How big data shape paediatric anaesthesia. Br J Anaesth. The Author(s); 2017;119:448–51.
3. Hoorweg A-LJ, Pasma W, van Wolfswinkel L, de Graaff JC. Incidence of Artifacts and Deviating Values in Research Data Obtained from an Anesthesia Information Management System in Children. Anesthesiology. 2018;128:293–304.
4. Kool NP, van Waes J a R, Bijker JB, Peelen LM, van Wolfswinkel L, de Graaff JC, et al. Artifacts in research data obtained from an anesthesia information and management system. Can J Anaesth. 2012;59:833–41.
5. Pasma W, Peelen LM, van Buuren S, van Klei WA, de Graaff JC. Artifact Processing Methods Influence on Intraoperative Hypotension Quantification and Outcome Effect Estimates. Anesthesiology. 2020;132:723–37.
6. Jakob S, Korhonen I, Ruokonen E, Virtanen T, Kogan A, Takala J. Detection of artifacts in monitored trends in intensive care. Comput Methods Programs Biomed. 2000;63:203–9.
7. Verduijn M, Peek N, de Keizer NF, van Lieshout EJ, de Pont ACJM, Schultz MJ, et al. Individual and Joint Expert Judgments as Reference Standards in Artifact Detection. J Am Med Informatics Assoc. 2008;15:227–34.
8. Du CH, Glick D, Tung A. Error-checking intraoperative arterial line blood pressures. J Clin Monit Comput. Springer Netherlands; 2019;33:407–12.
9. Gostt RK, Rathbone GD, Tucker AP. Real-time pulse oximetry artifact annotation on computerized anaesthetic records. J Clin Monit Comput. 2002;17:249–57.
10. Cunningham S, Symon AG, McIntosh N. The practical management of artifact in computerised physiological data. Int J Clin Monit Comput. 1994;11:211–6.
11. Friard O, Gamba M. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods Ecol Evol. 2016;7:1325–30.
12. Roth JA, Battegay M, Juchler F, Vogt JE, Widmer AF. Introduction to Machine Learning in Digital Healthcare Epidemiology. Infect Control Hosp Epidemiol. 2018;39:1457–62.

13. Toutenburg H. Fleiss, J. L.: Statistical Methods for Rates and Proportions. John Wiley & Sons, New York-London-Sydney-Toronto 1973. XIII, 233 S. Biom Z. 1974;16:539–539.
14. Kuhn M. Caret package. J Stat Softw. 2008;28.
15. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33:1–22.
16. Venables WN, Ripley BD. Modern applied statistics with S. 2002.
17. Schölkopf B, Smola AJ. Support Vector Machines and Kernel Algorithms. Handb Brain Theory Neural Networks. 2002;1119-25.
18. Settles B. Active Learning Literature Survey. Comput Sci Tech Report, 1648, Univ Wisconsin–Madison. 2009;
19. Chen L, Dubrawski A, Wang D, Fiterau M, Guillame-Bert M, Bose E, et al. Using Supervised Machine Learning to Classify Real Alerts and Artifact in Online Multisignal Vital Sign Monitoring Data. Crit Care Med. 2016;44:e456-63.


III

General discussion and summary

7

General discussion

Database research

The Chapters in this dissertation illustrate and discuss database research in the field of anesthesia. **Chapters 2** and **3** showed ways to re-use (large) anesthesia information management system (AIMS) databases for research. Despite being retrospective, such studies can still provide useful scientific information,[1] yet interpretation should always be done with caution.[2,3] **Chapters 4, 5** and **6** focused on one of many data quality aspects of an AIMS database, i.e. artifacts. AIMS database research and quality in general will be further discussed in the remainder of this chapter.

Database research is like making coffee, a complex sequence of actions with an interesting result. When we look at the making of a cup of coffee, the first step is selecting coffee beans from a place or multiple places in the world, which are roasted and further processed to hopefully get a delicious cup of black brew. The final product is however influenced by every step in between. It is important to understand that the quality of the coffee we drink cannot be any better than the quality of the original beans, only quality can be subtracted by imperfect processing. Factors that influence the quality are for example the age of the beans[4], the temperatures at which they are roasted[5,6], the time between the roast and making the coffee[7], the coarseness of the grinded coffee[8], the method of brewing, the temperature and the quality of the water that is used, the time the water was in contact with the coffee[9] and the temperature at which we drink coffee.[10] Each step in this process can also be viewed as a decision by the person who makes the coffee.

The same principle holds for database research. We cannot really add quality to the data which was stored; we can only extract what is already present in the data. Each step in the process of database research is a choice made by the researcher or the research team, which influences the quality of the end product. And just like in coffee, some of the decisions or processing steps are irreversible. In the process of database research, choices regarding data are often limited by data availability; typically only a selection of the source data is being used.[11] For example, when using only easily available and usable data a researcher runs the risk of ignoring other data of high quality that is hidden in for instance free texts.[12] Data is not immediately reused for scientific purposes. Because a certain quantity of study subjects is required, it can take several years before data is analyzed. Analog to

the storage of coffee also data has its expiration date. Using older data could for example lead to conclusions which are no longer relevant. Finally, in processing and analyzing the data, the researcher makes multiple decisions, each of which could bias estimates and influence the quality of science produced.

To illustrate the multitude of decisions a researcher has to make the following paragraphs describe the choices that were made in **Chapter 5** of this dissertation and discusses some of the alternatives and consequences. In **Chapter 5** we studied intraoperative hypotension and postoperative myocardial injury. We selected a subset of patients, at least 60 years of age who underwent medium to high risk surgery, and reoperations were excluded. In our center, these patients were per protocol all screened for postoperative myocardial injury in the three days after surgery using troponin measurements. Therefore the outcome information was complete. In other hospitals, where such a protocol is not in place a researcher might choose to assume that when troponin measurements were not available, that these particular patients did not have myocardial injury.[13] Another approach could be to include only those patients in the study in whom troponin values were determined. Such decisions could lead to information or selection bias. Besides using troponin levels as a biomarker that represents myocardial injury, a researcher could also decide to extract this outcome information from free text documentation in the electronic health record.

The determinant in our example, hypotension, also required several process decisions, which could all have affected the study results. We chose several ways to quantify hypotension, but typically only one definition for hypotension is used in research, which requires several extra decisions; Which blood pressure values should be used e.g. diastolic, mean or systolic blood pressure, invasive, non-invasive or both? Which measurements should be used e.g. all measurements during surgery, only post-induction or only a single measurement? Should measurements defined as artifacts, be removed? In that case, a method should be chosen to filter these artifacts. Which hypotension threshold should be used, e.g. an absolute threshold such as 65 mmHg or a relative threshold such as 20% below baseline?[14,15] Finally, which measure for hypotension should be chosen e.g. presence, duration or depth of hypotension or a combination of these, i.e. area under the threshold.[16]

As is obvious from the above example, in database research there is a high risk of decreasing quality during the fabrication of the end product; a scientific publication. The following paragraphs will go further into detail of some of the data quality issues, such as missing data and artifacts. Finally possible solutions are presented and the implications for database research and the research team are discussed.

Missing data

One of the important choices in database research is how to deal with missing data. For the development of the reference values in **Chapter 2**, we used commonly collected data: surgical specialty, ASA classification, height, weight, age and intraoperative blood pressure. After processing the data initially available in MPOG[17], we were only able to use two thirds of the data, because of missing data in sex, weight, height, ASA classification and surgical specialty. In this case complete case analysis was applied, which limited the cases available for the analysis.

When data could have been registered, but values are missing for a subset of patients, other decisions in the data processing chain can be made. In this respect, it is important to understand why certain data are missing. For example in **Chapter 3** we knew that when pre-anesthesia assessment forms are filled out by physicians, the questions which should be answered with 'no' were usually left unanswered and only relevant questions were answered with 'yes'. Hence, we considered all missing values to represent 'no', and completed the data as such before conducting the analyses. This does however require detailed information regarding the data registration process in clinical practice.

Another approach to missing data was used in the data used in **Chapter 5**. In the original study data from the AIMS and electronic health record were manually reviewed to correct for missing data.[18] In other words, different data sources were integrated (manually) to get a more complete database. In this case one should be aware that when gaps are filled manually, a mixed dataset is created with partly automatic and partly manual collected data. The quality of both types of data can be different. Statistical solutions to deal with missing data, such as multiple imputation, as we used in **Chapter 3**, can preserve the observed information [19,20]. A well substantiated decision on which of these methods should be applied has to be made based on the knowledge of data generating mechanism.

The effect of the decisions

All the above mentioned steps illustrate the complexity of database research. Each of the choices may affect the conclusion and quality of the eventual study results. Often however these effects are not systematically investigated, and hence unknown. This, together with sparse reporting of most of these processing decisions, makes studies harder to replicate or generalize.

The research field of intraoperative hypotension is an example where such comparative studies have been conducted. Systematic literature searches have identified a multitude of thresholds and measures that are used by researchers to express the presence and amount of intraoperative hypotension [14,21]. We have shown in **Chapter 5** that relative small decisions such as artifact removal methods also have a systematic effect on study outcomes. Larger effects were found when deciding which hypotension measure or threshold to use.[16] We used blood pressure data, in which we did not know which data points were artifacts. Therefore, without knowing the truth it is hard to measure the exact size of this bias. To better understand and quantify this effect, further research could focus on the simulation of artifacts and investigate the effect of these simulated artifacts on estimates.

Solving the quality issue

In **Chapter 4** we showed that anesthesia data used for research contains erroneous data. There are several ways to prevent the storage of such data. For example anesthesiologists could be asked to annotate or correct data during or after anesthesia care. This process could be enhanced by smart alerts from the system itself.[22] However, the anesthesiologist should focus on patient care, in which artifacts are not a big problem as they are easily recognized and ignored, rather than generate data perfectly suited for research. So there is a tradeoff between time spend on administration and the quality of the data. Configuration of the AIMS itself also could also have an effect on the quality of the data, e.g. interval of physiologic data storage, obligatory data fields, and enabling of editing of automatically captured data.[23] Choosing the latter could help data quality but could also have a detrimental effect on quality, similar to the smoothing of vital signs data in manual anesthesia records.[24]

A more durable and widely applicable solution would be to develop robust artifact filtering methods, which can identify and correct artefactual data. In **Chapter 5** we saw that this solution is not easy. Currently there is a wide selection of artifact filtering methods available in anesthesia care and research, but there is no consensus on which method should be used. Besides the variety in methods, the best method will depend on the situation in which data was collected. In **Chapter 6** we found that automatically identifying artifacts based on a learning algorithm does not yet work well enough to be broadly applicable. This finding might be due to the fact that information needed to correctly identify artifacts (as used by anesthesiologists in the setting of manual correction) is not always available during anesthesia or when data is eventually used retrospectively. Adding more information to the artifact algorithm might thus provide better accuracy.

But not every detail during anesthesia is registered in an AIMS, because this is not the main task of an anesthesiologist. A solution for this could perhaps be found in additional physiologic monitors and other sensors that can be connected with the AIMS software. For example, sensors that register the position of the operating table,[25] medicine administration parameters registered by infusion pumps,[26] or sensors registering door movement.[27] These may provide additional information to correctly identify artifacts, without interfering with anesthesia care. Taking this a step further, one could think of camera images from multiple angles and sound recordings, to be used to automatically detect events and actions, which are normally not (reliably) documented, e.g. surgical actions, patient movement or physiologic sensor manipulation.[27–30] This unconventional step would boost data quality and bring the context back to the data stored in AIMS.[31] Yet such a step requires careful ethical evaluation and supporting laws; physicians might fear that these data would also be queried for other purposes, such as liability lawsuits in case of calamities.

Nevertheless, the prospect of this amount of detail within the data provides endless possibilities for research and anesthesia care.[32] Better algorithms can be developed to further assist anesthesiologists not only with better documentation, but also with smarter alarm systems and eventually automating care itself with algorithms that take over certain tasks during anesthesia. This level of automation can be compared to aviation. However, just like in controlling airplanes, in health care we should be careful before implementing such algorithms. Algorithms which

assist pilots in flight are tested extensively in simulations, simulating numerous scenarios even extremely unlikely ones. In contrast, in health care algorithms are tested mostly by reviewing their performance in retrospective or prospective data, because detailed simulation of every type of patient is not (yet) possible.[33] Still, these data are inherently different than the "new" data in which the algorithm is active. For example over time the population of patients changes or doctors might change their registration and data generation habits, which causes concept drift making algorithms less accurate.[34] When algorithms are implemented, careful monitoring needs to be in place to observe prediction trends over time, model performance, model uncertainty, changes in patient outcome, adverse events and other unexpected anomalies.[33]

Even after extensive testing, unforeseen situations which were not simulated or experienced before could occur. Therefore pilots as well as doctors are required to always think for themselves and evaluate what is happening and whether the algorithm still makes sense. Trusting blindly on an algorithm in every situation or not understanding the inner workings of automatic decisions, can lead to disasters.[35] For example, the two crashes of Boeing 737 max planes in 2018 and 2019 were caused by a malfunctioning sensor on which automatic piloting algorithms were based. For commercial reasons, pilots were kept in the dark about the existence of a new system that adjusted the slopes of the plane. Regrettably, there was no documented way of turning off the system, and eventually these very skilled pilots lost their fights with the machine.[36] Anesthesiology and other health care processes are not yet that far automated as aviation. Nevertheless, advances in automatic systems such as closed loop systems for total intravenous anesthesia,[37] could suffer from similar issues, when anesthesiologists lose touch with the complex inner workings of such machines or black boxes. Clinicians should therefore be mindful for unexpected or unusual behavior and keep themselves up to date with any changes in these systems. Systems should stay transparent, and also inform its user when and how decisions were made.[35]

Implications

Research data generated from anesthesia databases is not perfect.[12,29,38–41] So, how will a researcher be able to confidently reuse anesthesia data, out of its context? When conducting research with anesthesia data, first and foremost it is important that one understands how the data was generated. For example when using physiological data, a researcher should have some idea on how physiologic patient monitors work; i.e. if there are already artifact filters in place and which values are stored in the database (for example the median or the current actual value). Also the researcher should be aware of the fact that there can still be errors in the data regardless of data preprocessing, consider why these occur or why certain values could be missing.

Second, conducting database research is team work. If a researcher does not have knowledge about the data, because data were received from another person, details should be inquired. Data specialists or data stewards in hospitals have a better overview of the resources to get an idea of the data generation mechanism. They coordinate the extraction, processing and sharing of the data.[42] If such a person is not available, which is not a rare phenomenon, the department should consider filling this vacancy.[11,43–46] Besides the data specialist, researchers also require help from clinicians to adequately interpret the data. Clinicians have the most information on how and why the anesthesia data was collected. Clinicians involved in database research should in their turn have adequate epidemiological and statistical knowledge to understand how the data will be analyzed and which issues could be expected. Nowadays, clinician should also have some understanding of other advanced techniques, such as machine learning and artificial intelligence, which are increasing in popularity.[32] Without the valuable input of every team member, wrong decisions can be made in the processing of data or faulty or irrelevant conclusions may be drawn from data.

Third, the knowledge about the data generation and each decision regarding data processing should be adequately reported. If the level of detail in which this is reported is limited, other researchers will be unable to reproduce the results. It also limits the insight of the reader about the magnitude of steps and assumptions required to get from the data to a scientific conclusion. Sharing the actual data and/or the statistical code along with the scientific paper might become more

important, when data and analysis get more complex. This method could over time also lead to standardization of data processing and analytics.[3]

Conclusions from studies based on database research should always be weighed carefully. The quality of the data, why they were documented and for which purpose they are suitable should be evaluated. If any scientific or clinical decision follows from the data, one should stay aware of the assumptions and decisions that have been used by a researcher or computer to come to this conclusion. Only then a final decision can be made supported by the data.

References

1. Simpao AF, Ahumada LM, Gálvez JA, Rehman MA: A review of analytics and clinical informatics in health care. J Med Syst 2014; 38
2. Choi SW, Wong GTC: Big data, big problems: causation, confounders and cohorts. Anaesthesia 2018; 73:384–7
3. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, Spreafico R, Hafler DA, McKinney EF: From big data to precision medicine. Front Med 2019; 6:1–14
4. Ross CF, Pecka K, Weller K: Effect of storage conditions on the sensory quality of ground arabica coffee. J Food Qual 2006; 29:596–606
5. Schenker S, Heinemann C, Huber M, Pompizzi R, Perren R, Escher F: Impact of roasting conditions on the formation of aroma compounds in coffee beans. J Food Sci 2002; 67:60–6
6. Baggenstoss J, Poisson L, Kaegi R, Perren R, Escher F: Coffee roasting and aroma formation: Application of different time-temperature conditions. J Agric Food Chem 2008; 56:5836–46
7. Glöss AN, Schönbächler B, Rast M, Deuber L, Yeretzian C: Freshness indices of roasted coffee: Monitoring the loss of freshness for single serve capsules and roasted whole beans in different packaging. Chimia (Aarau) 2014; 68:179-82
8. Cameron MI, Hofstetter D, Lee WT, Christopher H, Foster JM, Cameron MI, Morisco D, Hofstetter D, Uman E, Wilkinson J, Foster JM: Systematically Improving Espresso : Insights from Mathematical Modeling and Experiment. Matter 2019; 12:1–18
9. Spiro M, Selwood RM: The kinetics and mechanism of caffeine infusion from coffee: The effect of particle size. J Sci Food Agric 1984; 35:915–24
10. Cruz A, Green BG: Thermal stimulation of taste. Nature 2000; 403:889–92
11. Hofer IS, Levin MA, Simpao AF, McCormick PJ, Rothman BS: Anesthesia informatics grows up. Anesth Analg 2018; 127:18–20
12. Bezemer T, Groot MC de, Blasse E, Berg MJ Ten, Kappen TH, Bredenoord AL, Solinge WW van, Hoefer IE, Haitjema S: A Human(e) Factor in Clinical Decision Support Systems. J Med Internet Res 2019; 21:e11732
13. Salmasi V, Maheshwari K, Yang D, Mascha EJ, Singh A, Sessler DI, Kurz A: Relationship between Intraoperative Hypotension, Defined by Either Reduction from Baseline or Absolute Thresholds, and Acute Kidney and Myocardial Injury after Noncardiac Surgery: A Retrospective Cohort Analysis. Anesthesiology 2017; 126:47–65

14. Bijker JB, Klei WA van, Kappen TH, Wolfswinkel L van, Moons KGM, Kalkman CJ: Incidence of intraoperative hypotension as a function of the chosen definition: literature definitions applied to a retrospective cohort using automated data collection. Anesthesiology 2007; 107:213–20
15. Sessler DI, Bloomstone JA, Aronson S, Berry C, Gan TJ, Kellum JA, Plumb J, Mythen MG, Grocott MPW, Edwards MR, Miller TE, Grocott MP, Ackland GL, Brudney CS, Cecconi M, Ince C, Irwin MG, Lacey J, Pinsky MR, Sanders R, Hughes F, Bader A, Thompson A, Hoeft A, Williams D, Shaw AD, Aronson S, Kellum J, Bloomstone J, McEvoy MD, et al.: Perioperative Quality Initiative consensus statement on intraoperative blood pressure, risk and outcomes for elective surgery. Br J Anaesth 2019; 122:563–74
16. Vernooij LM, Klei WA van, Machina M, Pasma W, Beattie WS, Peelen LM: Different methods of modelling intraoperative hypotension and their association with postoperative complications in patients undergoing non-cardiac surgery. Br J Anaesth 2018; 120:1080–9
17. Kheterpal S: Clinical research using an information system: The multicenter perioperative outcomes group. Anesthesiol Clin 2011; 29:377–88
18. Waes JAR van, Nathoe HM, Graaff JC de, Kemperman H, Borst GJ de, Peelen LM, Klei WA van: Myocardial injury after noncardiac surgery and its association with short-term mortality. Circulation 2013; 127:2264–71
19. Donders ART, Heijden GJMG van der, Stijnen T, Moons KGM: Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006; 59:1087–91
20. Rubin DB: Multiple Imputation for Nonresponse in Surveys. Edited by Rubin DB. Hoboken, NJ, USA, John Wiley & Sons, Inc., 1987
21. Vernooij LM, Klei WA van, Moons KGM, Waes JA van, Peelen LM: Methods to express intraoperative hypotension exposure in the anaesthesia literature. Br J Anaesth 2020:72–5
22. Sandberg WS, Sandberg EH, Seim AR, Anupama S, Ehrenfeld JM, Spring SF, Walsh JL: Real-time checking of electronic anesthesia records for documentation errors and automatically text messaging clinicians improves quality of documentation. Anesth Analg 2008; 106:192–201
23. Epstein RH, Dexter F: Database quality and access issues relevant to research using anesthesia information management system data. Anesth Analg 2018; 127:105–14
24. Wax DB, Beilin Y, Hossain S, Lin HM, Reich DL: Manual editing of automatically recorded data in an anesthesia information management system. Anesthesiology 2008; 109:811–5

25. Hasegawa M, Nouri M, Fujisawa H, Hayashi Y, Inamasu J, Hirose Y, Yamashita J: Efficacy of monitoring patient's position during neurosurgical procedures: Introduction of real-time display and record. Neurol Med Chir (Tokyo) 2015; 55:305–10
26. Biltoft J, Finneman L: Clinical and financial effects of smart pump-electronic medical record interoperability at a hospital in a regional health system. Am J Heal Pharm 2018; 75:1064–8
27. Birgand G, Azevedo C, Rukly S, Pissard-Gibollet R, Toupet G, Timsit JF, Lucet JC: Motion-capture system to assess intraoperative staff movements and door openings: Impact on surrogates of the infectious risk in surgery. Infect Control Hosp Epidemiol 2019; 40:566–73
28. Vrigkas M, Nikou C, Kakadiaris IA: A review of human activity recognition methods. Front Robot AI 2015; 2:1–28
29. Simpao AF, Pruitt EY, Cook-Sather SD, Gurnaney HG, Rehman MA: The reliability of manual reporting of clinical events in an anesthesia information management system (AIMS). J Clin Monit Comput 2012; 26:437–9
30. Twinanda AP, Alkan EO, Gangi A, Mathelin M de, Padoy N: Data-driven spatio-temporal RGBD feature encoding for action recognition in operating rooms. Int J Comput Assist Radiol Surg 2015; 10:737–47
31. Levin MA, Wanderer JP, Ehrenfeld JM: Data, big data, and metadata in anesthesiology. Anesth Analg 2015; 121:1661–7
32. Connor CW: Artificial Intelligence and Machine Learning in Anesthesiology. Anesthesiology 2019; 131:1346–59
33. Khalifa M, Magrabi F, Gallego B: Developing a framework for evidence-based grading and assessment of predictive tools for clinical decision support. BMC Med Inform Decis Mak 2019; 19:1–17
34. Sebastião R, Silva MM, Rabiço R, Gama J, Mendonça T: Real-time algorithm for changes detection in depth of anesthesia signals. Evol Syst 2013; 4:3-12
35. Amalberti R: The paradoxes of almost totally safe transportation systems. Saf Sci 2001; 37:109–26
36. Langewiesche W: What Really Brought Down the Boeing 737 Max? 2018 at <https://www.nytimes.com/2019/09/18/magazine/boeing-737-max-crashes.html>
37. Dumont GA, Ansermino JM: Closed-loop control of anesthesia: A primer for anesthesiologists. Anesth Analg 2013; 117:1130–8

38. Kool NP, Waes J a R van, Bijker JB, Peelen LM, Wolfswinkel L van, Graaff JC de, Klei WA van: Artifacts in research data obtained from an anesthesia information and management system. Can J Anaesth 2012; 59:833–41
39. Niskanen M, Tuovinen T, Purhonen S, Vauhkonen S, Hendolin H: Quality assurance in anaesthetic practice: comparison between two methods in detecting complications. Acta Anaesthesiol Scand 2002; 46:896–901
40. Wrightson WAG: A comparison of electronic and handwritten anaesthetic records for completeness of information. Anaesth Intensive Care 2010; 38:1052–8
41. Palaniswamy S, Jain V, Chakrabarti D, Bharadwaj S, Sriganesh K: Completeness of manual data recording in the anaesthesia information management system: A retrospective audit of 1000 neurosurgical cases. Indian J Anaesth 2019; 63:797
42. Hartter J, Ryan SJ, MacKenzie CA, Parker JN, Strasser CA: Spatially Explicit Data: Stewardship and Ethical Challenges in Science. PLoS Biol 2013; 11
43. Poterack KA, Epstein RH, Dexter F: The anesthesiologist-informatician: A survey of physicians board-certified in both anesthesiology and clinical informatics. Anesth Analg 2018; 127:115–7
44. Giambrone GP, Hemmings HC, Sturm M, Fleischut PM: Information technology innovation: The power and perils of big data. Br J Anaesth 2015; 115:339–42
45. Lehmann CU, Gundlapalli A V., Williamson JJ, Fridsma DB, Hersh WR, Krousel-Wood M, Ondrula CJ, Munger B: Five Years of Clinical Informatics Board Certification for Physicians in the United States of America. Yearb Med Inform 2018; 27:237–42
46. Raghupathi W, Raghupathi V: Big data analytics in healthcare: promise and potential. Heal Inf Sci Syst 2014; 2:1–10

8

Summary

This dissertation describes two studies in which large anesthesia datasets were used (**Part I**). Retrospective large database research has a big potential, but also data quality issues exist that need to be addressed. In **Part II** we zoom in to one of these issues in anesthesia databases; i.e. artefacts in physiologic data.

Part I

In **Chapter 2** we have developed reference curves for blood pressure during anesthesia in children. We collected pediatric anesthesia data from 10 different hospitals in two different countries. With 116,362 anesthetic procedures we constructed sex specific references, adjusted for age, weight or height. The curves provided in **Chapter 2**, help clinicians with a better reference for intra-operative blood pressure monitoring compared to what was previously available. These references can also be used in research, as we illustrated in **Chapter 3**. We applied the references to a single-center pediatric cohort, to explore which children get a relatively low blood pressure during anesthesia. The blood pressure was adjusted for sex and height using the references from **Chapter 2**. We studied the association of patient and anesthesia procedure characteristics with this normalized blood pressure. Part of the variation in blood pressure could be explained by the procedure characteristics, such as application of a loco-regional technique. Only small effects of patient characteristics on blood pressure were estimated, therefore we could not define a typical child that develops low blood pressure during anesthesia. In **Chapter 3** we showed how the blood pressure references can be applied methodologically to normalize the pre-incision blood pressure. Future research could study the blood pressure during the entire procedure, instead of just the pre-incision period. Also other types of summary measures could be studied such as the area under the threshold, rather than the mean of the last three measurements before incision, which was used in **Chapter 3**. Another direction for future research could be to study the relation of the normalized blood pressure with postoperative outcome.

Part II

In **Part II** of this dissertation we focused on artifacts in physiologic measurements (i.e. heart rate, saturation, end tidal carbon dioxide, non-invasive blood pressure and invasive blood pressure) during anesthesia.

First, in **Chapter 4** we observed the incidence of artifacts in pediatric cases during 170 hours of anesthesia. Incidence of artifacts ranges from 0.5% for heart rate to 7.5% for end tidal carbon dioxide. We found that these artifacts did not occur at random and were dependent on different factors, e.g. type of measurement, age of the patient, type of surgery or the phase of surgery at which the measurement was done. As some of these factors could also be related to outcome, we hypothesized that artifacts or the way artifacts are filtered, could act as a confounder.

Second, in **Chapter 5** we tested this hypothesis by studying the effect of different artifact filtering methods on the result of an example study. The artifact filtering methods were identified with a systematic literature search. As an example we used hospital data from adults older than 60 who underwent medium to high risk surgery and analyzed the relation between hypotension and postoperative myocardial injury. We showed that there is indeed a small systematic effect of artifact filtering methods on the estimated relation between hypotension and myocardial injury.

Finally, in **Chapter 6** we focused in more detail on the artifacts in invasive blood pressure data and formulated a method to identify artifacts automatically. We showed that besides dependency on factors (e.g. type of surgery and phase of surgery) it also matters when and how artifacts are identified. When someone observes the procedure live, similar to **Chapter 4**, the artifact identification was different than when someone reviewed the data retrospectively, which is customary in database research. The information or context available live and retrospective is different, and therefore the conclusion drawn by an observer is different. Therefore we cannot simply use live or retrospective manual annotations as a golden standard for artifact identification. A clear definition of what is assumed to be an artifact should first be formulated and reported by a researcher. Despite the differences we hypothesized that the process of artifact identification could still be automated. In **Chapter 6** we applied different learning algorithms to model the identification of artifacts. In this study, the performance of these algorithms remained mediocre at best. Future research could focus on development of better performing algorithms using additional information and further optimization of the training of such algorithms to reduce the manual work required.


IV

Appendices

Appendix A: Summary in Dutch
Samenvatting in het Nederlands

Eerst was er pen en papier. De medische anesthesiedossiers werden met de hand geschreven door een chirurg of anesthesioloog in de operatiekamer. De eerste voorbeelden van anesthesiedossiers dateren uit de late 19e eeuw. Het doel van de gemaakte grafieken was om een beter overzicht te krijgen van het beloop van vitale functies van de patiënt tijdens de operatie en was niet ontworpen met het oog op onderzoek.

Om anesthesiedocumentatie te verbeteren begon men te experimenteren met computers die fysiologische gegevens van de bewakingsmonitor automatisch konden vastleggen. Op basis van dit principe werden vanaf eind jaren tachtig de eerste Anesthesie Informatie Management Systemen (AIMS) ontwikkeld en geïmplementeerd. Naast de fysiologische parameters konden medicatietoedieningen en andere gebeurtenissen elektronisch worden gedocumenteerd in het AIMS, om zodoende een volledig overzicht van de anesthesie te krijgen. Deze elektronische systemen zorgden voor een completer en beter dossier dan hun analoge voorgangers. Omdat de gegevens al elektronisch beschikbaar waren, werd het hergebruik van deze gegevens voor onderzoek veel gemakkelijker.

AIMS waren meestal standalone systemen, die niet communiceerden met andere medische dossiersystemen. Tegenwoordig kunnen deze systemen ingebed zijn in of verbonden zijn met het elektronische patiënten dossier (EPD), waarbij het complete digitale medische dossier beschikbaar komt voor anesthesist en onderzoeker. AIMS heeft niet alleen de administratie van anesthesie, maar heeft ook de wetenschappelijke output van anesthesieafdelingen over de hele wereld verbeterd, bijvoorbeeld op het gebied van bloeddrukmanagement en uitkomsten zoals postoperatief myocardiaal letsel en beroerte.

Dit proefschrift gaat over hergebruik van anesthesiedata voor onderzoek. Anesthesie databases zijn erg geschikt voor onderzoek, omdat zij rijke data bevatten van een duidelijk omschreven tijdspanne (anesthesie tijdens een chirurgische procedure). Binnen deze periode hebben de verzamelde gegevens een hoge dichtheid (metingen per minuut) en zijn andere gebeurtenisgegevens zoals medicatietoediening zeer nauwkeurig vastgelegd. Daarmee kan het verband tussen

gebeurtenissen tijdens anesthesie en eventueel nadelige effecten na anesthesie in detail worden bestudeerd. Dit proefschrift beschrijft in **Deel I** twee studies waarbij de hierboven beschreven anesthesiedata werden hergebruikt. Retrospectief "big data" onderzoek heeft grote potentie, maar kent ook uitdagingen wat betreft de datakwaliteit. **Deel II** van dit proefschrift verdiept zich in één van deze kwesties, namelijk meetfouten (artefacten) in fysiologische gegevens, zoals bloeddruk en hartslag.

Deel I

In **Hoofdstuk 2** hebben we referentiecurves ontwikkeld voor bloeddruk tijdens anesthesie bij kinderen. We hebben gegevens verzameld van kinderen onder anesthesie uit 10 verschillende ziekenhuizen in twee verschillende landen. Met de data van 116.362 anesthesieprocedures hebben we bloeddrukreferentiecurves opgesteld voor jongens en meisjes, gecorrigeerd voor leeftijd, gewicht of lengte. Met deze curves kunnen clinici beter de intra-operatieve bloeddruk beoordelen dan met wat eerder beschikbaar was. De referenties kunnen daarnaast worden gebruikt voor wetenschappelijk onderzoek. Dit hebben we geïllustreerd in **hoofdstuk 3** waar we de referenties hebben toegepast op de data van een cohort van één centrum. Vervolgens hebben we bekeken welke kinderen een relatief lage bloeddruk krijgen tijdens anesthesie. Met behulp van de referenties uit **Hoofdstuk 2** is de bloeddruk genormaliseerd voor geslacht en lengte. We onderzochten de associatie van patiënt en anesthesie-eigenschappen met de genormaliseerde bloeddruk. Een deel van de variatie van de bloeddruk kon worden verklaard door procedurekenmerken, zoals toepassing van een locoregionale techniek. Er werden slechts kleine effecten van patiëntkenmerken op de bloeddruk gevonden, daarom konden we niet een type kind beschrijven dat een lage bloeddruk ontwikkelt tijdens anesthesie. In **Hoofdstuk 3** toonden we echter wel aan hoe de bloeddrukreferenties methodologisch kunnen worden toegepast door het normaliseren van de bloeddruk. We hebben in dit onderzoek alleen gekeken naar de bloeddruk voor incisie. Toekomstig onderzoek zou zich kunnen focussen op de bloeddruk tijdens de hele procedure. Daarnaast zouden andere samenvattende maten voor de intra-operatieve bloeddruk kunnen worden bestudeerd, zoals de totaal gemeten verlaagde bloeddruk ("area under the threshold"). Verder zou kunnen worden onderzocht wat de relatie is tussen de genormaliseerde bloeddruk en uitkomsten na anesthesie.

Deel II

In **Deel II** van dit proefschrift hebben we ons gericht op meetfouten ofwel artefacten in fysiologische metingen tijdens anesthesie die een probleem zouden kunnen vormen bij het uitvoeren van wetenschappelijk onderzoek met deze data. De onderzochte parameters zijn hartslag, saturatie, "end-tidal CO2", non-invasieve bloeddruk en invasieve bloeddruk.

Allereerst hebben we in **Hoofdstuk 4** de incidentie van artefacten in pediatrische anesthesie gemeten door 170 uur anesthesie te observeren. De incidentie van artefacten varieerde van 0,5% voor de hartslag tot 7,5% voor "end-tidal CO2". We ontdekten dat deze artefacten niet willekeurig voorkwamen in de data en onder andere afhankelijk waren van het type meting, de leeftijd van de patiënt, het type operatie of de fase van de operatie waarbij de meting werd uitgevoerd. Aangezien sommige van deze factoren ook verband zouden kunnen hebben met klinische uitkomsten, kwamen we tot de hypothese dat deze artefacten ofwel de manier waarop ze uit data worden gefilterd, zouden kunnen zorgen voor systematische fouten (bias) in klinisch wetenschappelijk onderzoek.

Om deze hypothese te testen hebben we in **Hoofdstuk 5** het effect onderzocht van verschillende artefactfiltermethoden op de conclusie van een voorbeeldstudie. De toegepaste filtermethoden hebben we aan de hand van een systematisch literatuuronderzoek geïdentificeerd. Als voorbeeld zijn de data gebruikt van volwassenen die ouder zijn dan 60 en chirurgie ondergingen met een gemiddeld tot hoog risico. We analyseerden in het cohort de relatie tussen hypotensie (verlaagde bloeddruk) en postoperatieve myocardschade. We toonden aan dat er inderdaad een klein systematisch effect is van artefactfiltermethoden op de geschatte relatie tussen hypotensie en myocardiaal letsel.

Ten slotte hebben we ons in **Hoofdstuk 6** verder verdiept in de artefacten in invasieve bloeddrukdata en onderzocht of artefacten automatisch geïdentificeerd kunnen worden. We hebben laten zien dat naast afhankelijkheid van factoren (zoals type van chirurgie en fase van de operatie) het er ook toe doet wanneer en hoe artefacten worden geïdentificeerd. Wanneer iemand tijdens de procedure aanwezig was om artefacten te observeren (zoals in **Hoofdstuk 4),** werden andere artefacten geïdentificeerd dan wanneer iemand de gegevens achteraf beoordeelde op artefacten, wat gebruikelijk is in databaseonderzoek. De informatie of context

die live en retrospectief beschikbaar is verschilt en daarom is de conclusie van een waarnemer anders. Daarom kunnen deze registraties niet zomaar worden gebruikt als referentie voor het ontwikkelen van een algoritme dat artefacten identificeert. Het is belangrijk dat een duidelijke definitie van wat onder een artefact wordt verstaan geformuleerd en gerapporteerd wordt door de onderzoeker. Ondanks de verschillen in definitie veronderstelden we dat artefactidentificatie kan worden geautomatiseerd op een uniforme wijze. In **Hoofdstuk 6** hebben we verschillende "machine learning" modellen toegepast voor de identificatie van artefacten. In deze studie waren de prestaties van deze algoritmen op zijn hoogst middelmatig. Daarom zou toekomstig onderzoek zich kunnen focussen op de ontwikkeling van beter presterende algoritmen met bijvoorbeeld aanvullende informatie en verdere optimalisatie van het leren van dergelijke algoritmen om zodoende het benodigde handmatige werk te verminderen.

Appendix C: About the author

Curriculum Vitae

Wietze Pasma was born on the 20th of April 1982 in Hengelo (ov). After graduating from high school in 2000 (Bataafse Kamp, Hengelo (ov)) he started his study of Veterinary Medicine at Utrecht University. In 2008 he graduated in small animal health. After working as a veterinarian for a short period he started working in the UMC Utrecht in 2011 as a datamanager. He setup datamanagement for anesthesiology, intensive care and acute healthcare. In this role he successfully combined healthcare knowledge with IT knowledge. He initiated the UMC Utrecht data contribution to MPOG and worked on several research projects on the background.

In 2015 he got the opportunity to start his own research as a PhD candidate at the Graduate School of Life Sciences (Utrecht University) supervised by Prof. dr. Wilton van Klei, Prof. dr. Stef van Buuren, dr. Jurgen de Graaff and dr. Linda Peelen. From 2015 up to 2020 he combined his PhD research with work as a datamanager.

Appendix D: List of publications

This dissertation

Pasma W, Graaff JC de, Buuren S van, Duijghuisen JJ, Nafiu OO, Kheterpal S, Klei WA van: Reference Values for Noninvasive Blood Pressure in Children during Anesthesia: A Multicentered Retrospective Observational Cohort Study. Anesthesiology 2016; 125:904–13

Pasma W, Peelen LM, Broek S van den, Buuren S van, Klei WA van, Graaff JC de: Patient and anesthesia characteristics of children with low pre-incision blood pressure: A retrospective observational study. Acta Anaesthesiol Scand 2020; 64:472–80

Hoorweg A-LJ, **Pasma W**, Wolfswinkel L van, Graaff JC de: Incidence of Artifacts and Deviating Values in Research Data Obtained from an Anesthesia Information Management System in Children. Anesthesiology 2018; 128:293–304

Pasma W, Peelen LM, Buuren S van, Klei WA van, Graaff JC de: Artifact Processing Methods Influence on Intraoperative Hypotension Quantification and Outcome Effect Estimates. Anesthesiology 2020; 132:723–37

Pasma W, Wesselink EM, Buuren S van, Graaff JC de, Klei WA van: Artifacts annotations in anesthesia blood pressure data by man and machine. 2020 Accepted for publication in Journal of Clinical Computing and Monitoring

Other

Wolters AE, Dijk D van, **Pasma W**, Cremer OL, Looije MF, Lange DW de, Veldhuijzen DS, Slooter AJC: Long-term outcome of delirium during intensive care unit stay in survivors of critical illness: a prospective cohort study. Crit Care 2014; 18:R125

Bruin L de, **Pasma W**, Werff DBM van der, Schouten TANJ, Haas F, Zee DC van der, Wolfswinkel L van, Graaff JC de: Perioperative hospital mortality at a tertiary paediatric institution. Br J Anaesth 2015; 115:608–15

Soliman IW, Lange DW de, Peelen LM, Cremer OL, Slooter AJC, **Pasma W**, Kesecioglu J, Dijk D van: Single-center large-cohort study into quality of life in Dutch intensive care unit subgroups, 1 year after admission, using EuroQoL EQ-6D-3L. J Crit Care 2015; 30:181–6

Koster-Brouwer ME, Groep K van de, **Pasma W**, Smeets HM, Slooter AJC, Lange DW de, Dijk D van, Poll T van der, Bonten MJM, Cremer OL, MARS Consortium: Chronic healthcare expenditure in survivors of sepsis in the intensive care unit. Intensive Care Med 2016; 42:1641–2

Wolters AE, Peelen LM, Veldhuijzen DS, Zaal IJ, Lange DW de, **Pasma W**, Dijk D van, Cremer OL, Slooter AJC: Long-Term Self-Reported Cognitive Problems After Delirium in the Intensive Care Unit and the Effect of Systemic Inflammation. J Am Geriatr Soc 2017; 65:786–91

Klei WA van, Waes JAR van, **Pasma W**, Kappen TH, Wolfswinkel L van, Peelen LM, Kalkman CJ: Relationship Between Preoperative Evaluation Blood Pressure and Preinduction Blood Pressure: A Cohort Study in Patients Undergoing General Anesthesia. Anesth Analg 2017; 124:431–7

Vernooij LM, Klei WA van, Machina M, **Pasma W**, Beattie WS, Peelen LM: Different methods of modelling intraoperative hypotension and their association with postoperative complications in patients undergoing non-cardiac surgery. Br J Anaesth 2018; 120:1080–9

Pancaro C, Shah N, **Pasma W**, Saager L, Cassidy R, Klei W van, Kooij F, Vittali D, Hollmann MW, Kheterpal S, Lirk P: Risk of Major Complications After Perioperative Norepinephrine Infusion Through Peripheral Intravenous Lines in a Multicenter Study. Anesth Analg 2019