Big Data in Neonatal Intensive Care

Carolyn McGregor
University of Ontario Institute of Technology, Canada

The effective use of big data within neonatal intensive care units has great potential to support a new wave of clinical discovery, leading to earlier detection and prevention of a wide range of deadly medical conditions.

Premature or preterm birth, defined as birth before a gestational age of 37 weeks, constitutes one of the most significant perinatal health problems in industrialized nations. In Canada, for example, preterm birth accounts for 75 to 85 percent of all perinatal mortality—that is, stillbirths plus deaths of infants less than 7 days old.[1,2]

Neonatal intensive care units (NICUs) provide critical care for premature and ill full-term infants. Critical care broadly utilizes significant resources. In the US and Canada, the cost of critical care across all ages is between 20 and 30 percent of total healthcare organization expenditures.[3] In Canada, neonatal patients represent approximately 15 percent of all critical care admissions; because the average length of stay is almost three times as long as that for all other critical care cohorts, these patients constitute a major critical care group.[4]

Premature and ill full-term infants require complex, real-time, clinical decision support correlating medical data from multiple sources as there are many potential complications of prematurity such as infection or damage to developing brain, lungs, or eyes. As the infant ages, the dangers can diminish if complications have not arisen as the organs develop and the infant grows larger and stronger. However, premature and ill full-term infants discharged from the NICU can suffer lifelong problems such as lung damage or blindness, or demonstrate neurological issues as they grow such as attention deficit-hyperactivity disorder.

Neonatal intensive care is a complex environment that must support collaborative decision making among various care providers. Despite the ongoing transition from paper to electronic charting, providers continue to discuss much clinical information qualitatively, such as "this baby is having spells," rather than quantitatively due to the lack of tools, techniques, and policies to support the analysis of complex, high-frequency, physiological data streams.[5]

This "big data" remains an untapped resource that has the potential to significantly improve medical discovery, quality of care, and productivity. Clinical decision support in neonatal intensive care—and critical care generally—would benefit from online health analytics platforms that leverage high-speed physiological data together with other electronic health record data. My research team has developed one such platform, Artemis, to meet this need.

NEONATAL INTENSIVE CARE: A BIG DATA PROBLEM

As Figure 1 shows, a range of medical devices monitor the vital organs of neonatal infants within NICUs. These patients might also be assisted by mechanical ventilation or have drugs or other fluids administered via infusion pumps. Monitors attached to these devices

Published by the IEEE Computer Society

display constantly changing sensed data, usually at second-by-second intervals, that a caregiver—typically a nurse—must translate into actionable information. The devices issue audible and visual alerts when readings breach standardized-population-based thresholds, indicating a significant risk to the patients' health and the stability of their vital organs.

Many NICU patients have heart activity monitored by electrocardiography (ECG), which can sample up to 1,000 readings a second to construct a waveform signal demonstrating the functioning of the heart. This translates to 86.4 million readings a day per patient. From this source signal, the ECG device also derives the heart rate and respiration rate, with each of these signals producing 86,400 readings a day per patient. Another data waveform, derived from the impedance between each of the three ECG leads attached to the patient's chest, measures chest wall movement to assess breathing rates. This signal can result in 5.4 million data points a day per patient.

A newborn's neurological function can also be monitored, resulting in multiple waveforms that each generate tens of millions of data points per patient per day.

Blood oxygen saturation readings are usually taken from a sensor on the patient's foot. From the source signal, the monitor can derive a second-by-second pulse rate, resulting in 86,400 readings per patient per day. This can be used together with or as an alternative to the ECG-derived heart rate if the latter is unavailable or occluded by artifacts.[6]

Figure 1. **Neonatal patients receive extensive physiological and environmental monitoring. (Top: US Air Force photo/Staff Sgt. Matthew Rosine; bottom: photo by Bobjgalindo.)**

Drug and nutrition infusion data from smart infusion pumps (SIPs) contribute to the big data problem. SIPs can provide more than 60 different types of data every 10 seconds. If a newborn stays in the NICU for 30 days, one SIP can generate 4.4 Mbytes of data per hour, 106 Mbytes of data per day, and 3 Gbytes of data monthly. Preterm infants can be connected to up to 13 SIPs, resulting in 39 Gbytes of drug infusion data from a single patient per month.[7]

Due to limited capital resources, medical monitoring devices within NICUs can be up to 15 years old. However, even older monitors have the capacity to output real-time, high-frequency data to clinical decision support systems.

CURRENT NICU DISCOVERY CHALLENGES

Clinical decision makers currently are not able to leverage the orders of magnitude of data being generated in NICUs. Typical medical devices have a rolling memory of approximately 72 hours to accommodate the Monday review of Friday and weekend data, but traditional charting protocols do not facilitate the use of that data.

For example, a premature newborn's heart beats approximately 7,000 times an hour. Yet, both paper and electronic charting protocols enable the persistent storage of only one value per hour of an indicative heart rate.

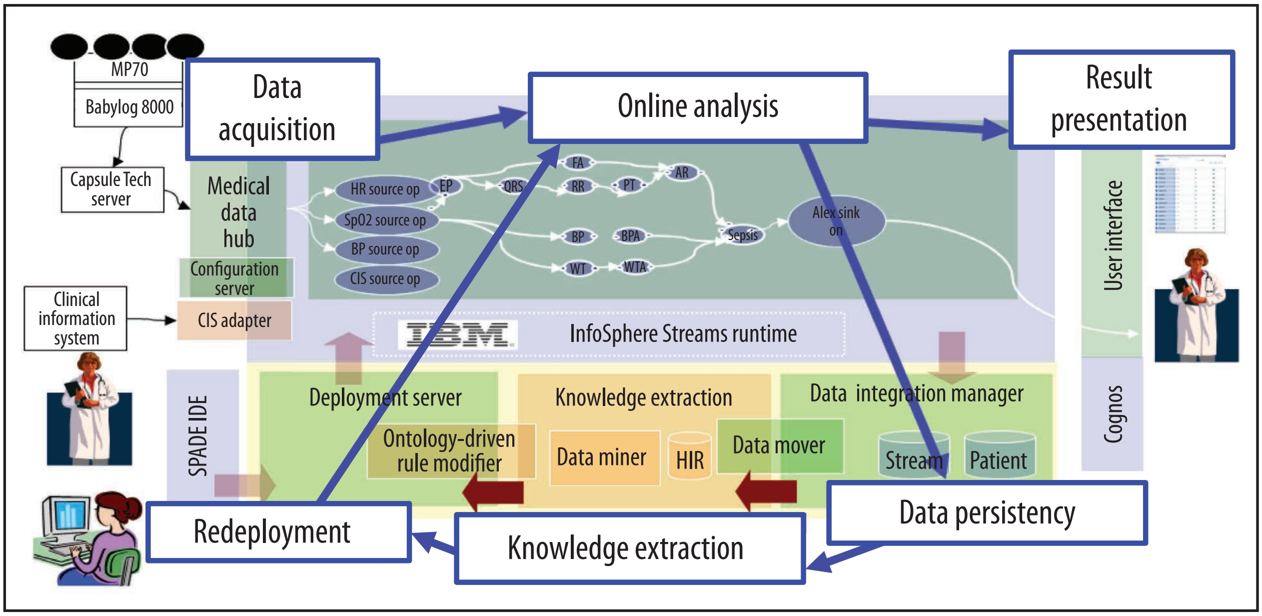

Figure 2. Artemis platform. Artemis enables concurrent diagnoses of multiple patients through real-time analysis of multiple data streams.

Caregivers employ qualitative rather than quantitative heuristics to determine the number to record, as part of the objective is to express overall stability or instability hour to hour. However, second-by-second data more accurately captures retinal exposure to oxygen than top-of-the-hour spot readings of blood oxygen saturation.[8]

Within the past 15 years, medical studies have demonstrated the potential for physiological data streams to reveal early signs of the onset of a small set of devastating conditions within the neonatal population, including late-onset neonatal sepsis (a form of hospital-acquired infection),[9] pneumothorax (collapsed lung),[10] intraventricular hemorrhage (bleeding in the brain),[11,12] and periventricular leukomalacia (death of brain tissue).[13] However, due to the limited nature of the black-box medical devices used to extract, store, and analyze such data, these studies could only examine a particular type of data, such as ECG signals, or clinical condition.[14,15]

Opportunities abound to study many other conditions, but more robust big data infrastructures are needed. Because many clinical studies involve multiple institutions, these infrastructures must span organizational boundaries.

Current use of information technology within NICUs does not support the discovery of new patterns in high-frequency physiological data streams that could lead to earlier disease detection and reduced mortality and morbidity, as well as tools to rapidly translate clinical research results to new treatment guidelines and bedside practices.

A new platform is thus required for both real-time and retrospective analysis of physiological data from multiple data streams, to handle multiple conditions and diseases of multiple patients at multiple locations. At the same time, clinical practice and healthcare professional training guidelines must evolve to support the use of such a platform.

ARTEMIS: CLINICAL DECISION SUPPORT FOR CRITICAL CARE

Artemis is an online health analytics platform that enables concurrent diagnoses of multiple patients through real-time analysis of multiple data streams.[15,16] Figure 2 shows the general architecture of the platform,[15] which is named after the Greek goddess of childbearing due to the initial NICU context for case study research.

The platform's *data acquisition* component continuously inputs physiological data streams from medical devices together with available clinical information. It then forwards this data to the *online analysis* component, which employs IBM's InfoSphere Streams middleware system to process the data in real time. Artemis stores the original data, together with the newly generated analytics, in the *data persistency* component. The *knowledge extraction* component mines data specifically tailored to support clinical research for a range of conditions. The *redeployment* component then feeds new clinically validated algorithms to the online analysis component. Artemis processes and stores both raw and derived data from multiple infants at the rate it is generated.[15,16]

Artemis was initially deployed in the NICU at The Hospital for Sick Children (SickKids) in Toronto in August

2009 with physiological data acquired from Philips IntelliVue MP70 neonatal monitors and relevant clinical information. The SickKids NICU monitors all neonatal patients using ECG and derived signals such as heart rate, respiration rate, and chest impedance for breath detection together with blood oxygen saturation. It also collects systolic, diastolic, and mean blood pressure data when available. The platform's data acquisition and online analysis components are located at SickKids, while the data persistency, knowledge extraction, and redeployment components reside at the University of Ontario Institute of Technology (UOIT).[16]

To date, Artemis has collected data at SickKids for more than 400 neonatal patients. This environment supports clinical studies on

- late-onset neonatal sepsis;[16,17]
- apnea of prematurity, in which the infant experiences pauses in breathing and reductions in heart rate and blood oxygen saturation;[5]
- retinopathy of prematurity, which can result in permanent blindness;[8] and
- pain.[18]

In April 2010, a new version of Artemis, Artemis Cloud, began collecting data from the Women and Infants Hospital of Rhode Island (WIHRI). This cloud-based environment takes spot readings each minute from bedside Spacelabs Healthcare monitoring devices and transmits these as HL7 packets through a secure Internet tunnel to UOIT. More than 250 WIHRI patients have been enrolled thus far in a study on neonatal instability.[15,17]

A third Artemis installation that contains only the data persistency, knowledge extraction, and redeployment components was recently implemented to enable retrospective data mining on a dataset of nearly two years of 30-second spot readings obtained from 1,151 patients in the SickKids NICU to refine clinical decision making for earlier detection of late onset neonatal sepsis, ultimately through the redeployment component.[9]

The current Artemis implementations at SickKids and WIHRI have no impact on bedside care yet. We are comparing analytical results with current clinical observation and treatment practices to discover new patterns in real-time physiological data that could lead to the earlier detection and prevention of various diseases.[5,15]

IMPACT ON RESEARCH, EDUCATION, AND POLICY

Real-time analysis of high-frequency physiological data could be as positively disruptive to healthcare as genomics research. However, a lack of medical informatics research has hampered progress in this area. More work is required to

- extract, securely transmit, process, and store electronic data streams together with any derived features and abstractions;
- support retrospective and prospective clinical studies using generalized systemic approaches such as Artemis;
- provide effective and appropriate visualizations for the various critical care roles;
- support evaluation of new discoveries through clinical trials;
- develop mechanisms to translate extracted knowledge to the patient bedside without introducing another black box;
- support quality-improvement initiatives to clinical treatment and training guidelines; and
- provide tools to easily measure and demonstrate improved healthcare service through the adoption of these new clinical guidelines.

Real-time analysis of high-frequency physiological data could be as positively disruptive to healthcare as genomics research.

The resulting technologies, tools, and methods must be commercialized for use within healthcare organizations.

Beyond the need for systemic research and commercialization, healthcare workers must be trained in the use of such systems. However, engaging skilled workers to participate in healthcare informatics projects has been hampered by the industry's failure to formally recognize informatics as a specialty or subspecialty.

This is changing, however. In November 2008, the American Medical Informatics Association (AMIA) recommended new core content and training requirements[19] to "transform healthcare by analyzing, designing, implementing, and evaluating information and communication systems that enhance individual and population health outcomes, improve patient care, and strengthen the clinician-patient relationship."[20] Clinical informaticians should combine their knowledge of patient care and informatics concepts, methods, and tools to

- "assess information and knowledge needs of healthcare professionals and patients;
- characterize, evaluate, and refine clinical processes;
- develop, implement, and refine clinical decision-support systems; and
- lead or participate in the procurement, customization, development, implementation, management, evaluation, and continuous improvement of clinical information systems."[20]

AMIA's decision was a milestone in recognizing the importance of informatics in healthcare, and will hopefully lead to international adoption of a clinical informatics subspecialty in various medical disciplines. In principle, AMIA's core content and training recommendations apply to all types of critical care, but guidelines on the use of high-speed physiological data and analytical tools and techniques must be tailored to each subspecialty.

The regulatory landscape for clinical decision-support technologies in many countries has constantly changed during the past decade as such technologies have evolved. New regulatory approaches are needed that embrace a multiplatform paradigm such as Artemis instead of the traditional "box at the bedside" paradigm, which, due to tight capital expenditure budgets, prevents hospitals from implementing improvements in a timely fashion.

System developers and policymakers must also develop guidelines to anonymize streaming physiological data. A recent, yet-to-be-published survey of Australian and Canadian citizens we conducted found strong support for the secondary use of such data for medical research, but only with appropriate security and privacy-protection mechanisms in place.

The effective use of big data within neonatal intensive care units has great potential to support a new wave of clinical discovery, leading to earlier detection and prevention of a wide range of deadly medical conditions. The ability to process multiple high-speed physiological data streams from multiple patients in multiple locations and in real time could dramatically improve both healthcare efficiency and patient outcomes.

To realize this vision, researchers must develop standardized, robust, online health analytics platforms similar to Artemis as well as quality improvement processes that translate analytical results into new clinical guidelines. In addition, caregivers and IT personnel operating such platforms must be trained in the use of big data tools and techniques. This in turn will require new educational programs within the domain of health informatics.

While the focus here is on NICUs, big data within other critical care units can likewise be leveraged. Another promising medical application of big data is in remote monitoring, which is used by many ambulatory and chronic-care patients.

Given the broad social implications of the use of sensor data, system developers and policymakers must work together to ensure that patients' privacy and confidentiality is protected when their data is used for general research purposes. C

References

1. M.S. Kramer et al., "Secular Trends in Preterm Birth: A Hospital-Based Cohort Study," *J. Am. Medical Assoc.*, vol. 280, no. 21, 1998, pp. 1849-1854.
2 D.F. Smith, "Measuring Up: A Health Surveillance Update on Canadian Children and Youth," *Paediatrics & Child Health*, vol. 4, no. 8, 1999, pp. 551-552.
3. P. Jacobs and T.W. Noseworthy, "National Estimates of Intensive Care Utilization and Costs: Canada and the United States," *Critical Care Medicine*, vol. 18, no. 11, 1990, pp. 1282-1286.
4. K. Lee et al., "CIHI Survey: Intensive Care in Canada," *Healthcare Q.*, vol. 9, no. 1, 2006, pp. 32-33.
5. A. Thommandram et al., "Classifying Neonatal Spells Using Real-Time Temporal Analysis of Physiological Data Streams: Algorithm Development," *Proc. IEEE EMBS Special Topic Conf. Point-of-Care Healthcare Technologies* (PHT 13), IEEE, 2013, pp. 240-243.
6. J.M. Feldman, M.H. Ebrahim, and I. Bar-Kana, "Robust Sensor Fusion Improves Heart Rate Estimation: Clinical Evaluation," *J. Clinical Monitoring*, vol. 13, no. 6, 1997, pp. 379-384.
7. N. Bressan, A. James, and C. McGregor, "Integration of Drug Infusion Data with Physiological Data Streams Using a Cloud Computing Paradigm," to appear in *Proc. 35th Ann. Int'l Conf. IEEE Eng. in Medicine & Biology Society* (EMBS 13), IEEE, 2013.
8. J. Cirelli et al., "Analysis of Continuous Oxygen Saturation Data for Accurate Representation of Retinal Exposure to Oxygen in the Preterm Infant," *Enabling Health and Healthcare through ICT*, K.L. Courtney, O. Shabestari, and A. Kuo, eds., IOS Press, 2013, pp. 126-131.
9. C. McGregor, C. Catley, and A. James, "Variability Analysis with Analytics Applied to Physiological Data Streams from the Neonatal Intensive Care Unit," *Proc. 25th IEEE Int'l Symp. Computer-Based Medical Systems* (CBMS 12), IEEE, 2012; doi:10.1109/CBMS.2012.6266385.
10. N. McIntosh et al., "Clinical Diagnosis of Pneumothorax Is Late: Use of Trend Data and Decision Support Might Allow Preclinical Detection," *Pediatric Research*, vol. 48, no. 3, 2000, pp. 408-415.
11. J. Fabres et al., "Both Extremes of Arterial Carbon Dioxide Pressure and the Magnitude of Fluctuations in Arterial Carbon Dioxide Pressure Are Associated with Severe Intraventricular Hemorrhage in Preterm Infants," *Pediatrics*, vol. 119, no. 2, 2007, pp. 299-305.
12. V. Tuzcu et al., "Altered Heart Rhythm Dynamics in Very Low Birth Weight Infants with Impending Intraventricular Hemorrhage," *Pediatrics*, vol. 123, no. 3, 2009, pp. 810-815.
13. S. Shankaran et al., "Cumulative Index of Exposure to Hypocarbia and Hyperoxia as Risk Factors for Periventricular Leukomalacia in Low Birth Weight Infants," *Pediatrics*, vol. 118, no. 4, 2006, pp. 1654-1659.
14. M. Stacey and C. McGregor, "Temporal Abstraction in Intelligent Clinical Data Analysis: A Survey," *Artificial Intelligence in Medicine*, vol. 39, no. 1, 2007, pp. 1-24.
15. C. McGregor et al., "Next Generation Neonatal Health Informatics with Artemis," *User Centered Networked Health Care*, A. Moen et al., eds., IOS Press, 2011, pp. 115-119.
16. M. Blount et al., "Real-Time Analysis for Intensive Care: Development and Deployment of the Artemis Analytic System," *IEEE Eng. in Medicine and Biology Magazine*, vol. 29, no. 2, 2010, pp. 110-118.

17. C. McGregor et al., “Late Onset Neonatal Sepsis Detection in Newborn Infants via Multiple Physiological Streams,” *J. Critical Care*, vol. 28, no. 1, 2013, pp. e11-e12.
18. T. Naik et al., “Design of Temporal Analysis for a Novel Premature Infant Pain Profile Using Artemis,” *J. Critical Care*, vol. 28, no. 1, 2013, p. e4.
19. D.E. Detmer, J.R. Lumpkin, and J.J. Williamson, “Defining the Medical Subspecialty of Clinical Informatics,” *J. Am. Medical Informatics Assoc.*, vol. 16, no. 2, 2009, pp. 167-168.
20. R.M. Gardner et al., “Core Content for the Subspecialty of Clinical Informatics,” *J. Am. Medical Informatics Assoc.*, vol. 16, no. 2, 2009, pp. 153-157.

Carolyn McGregor *is the Canada Research Chair in Health Informatics and a professor at the University of Ontario Institute of Technology, Canada. Her research focuses on critical care health informatics, in particular on the use of big data in neonatal health informatics. McGregor received a PhD in computer science from the University of Technology, Sydney, Australia. She is a senior member of IEEE. Contact her at c.mcgregor@ieee.org.*