SPECIAL ARTICLE

From computer to bedside, involving neonatologists in artificial intelligence models for neonatal medicine

Daniel Vijlbrief [1]✉, Jeroen Dudink[1], Wouter van Solinge[2], Manon Benders[1] and Saskia Haitjema[2]

In recent years, data have become the main driver of medical innovation. With increased availability and decreased price of storage and computing power, the potential for improvement in care is enormous. Many data-driven explorations have started. However, the actual implementation of artificial intelligence in healthcare remains scarce. We describe essential elements during a computer-to-bedside process in a data science project that support the crucial role of the neonatologist.

IMPACT:

- There is a great potential for data science in neonatal medicine.
- Multidisciplinary teams form the foundation of a data science project.
- Domain experts will need to play a pivotal role.
- We need an open learning environment.

INTRODUCTION

Recent advances in digital data collection in electronic health record (EHR) systems and data processing using computer analytics paved the way for data-driven innovation using artificial intelligence (AI) in healthcare. Digital health systems employing AI algorithms can now be developed to monitor patients and support clinicians in making diagnostic and therapeutic decisions.[1]

However, few AI algorithms have been implemented in clinical care.[1] Several reasons have been identified, including the European regulatory framework, General Data Protection Regulation, and Medical Device Regulation, which is currently unclear on dealing with self-learning. Yet, we and others undertook efforts to implement algorithms in clinical care.[2,3] Despite good intentions of researchers, there are challenges. Many AI models are developed as a research paper, and the primary objective is optimizing model performance. They subsequently fail to validate, calibrate, and incorporate the model into an application to bridge the gap between the computer and the bedside.

We have recently gained experience developing and implementing digital health solutions within a corporate applied data analytics program. The Applied Data Analytics in Medicine program was initiated in 2017 by the UMC Utrecht.[4] Its foundations were the data analytics value chain, the innovation funnel, and the multidisciplinary project team. The analytics value chain communicates the basic concept that an actionable insight needs to be generated to create value (Fig. 1). The funnel supports controlled innovation by the definition of go/no-go gates in between the development and implementation phases (Fig. 2). We strongly feel that medical needs should be the starting point

of every project, with the healthcare professional as a product owner.[5] This is, however, not without challenges. We aim to highlight four focal points that are often overlooked and where neonatologists can immediately play a key role.

MULTIDISCIPLINARY WAY OF WORKING

Clinical data science needs a multidisciplinary team built around three group representatives, a healthcare professional (domain expert), a patient representative, and a data scientist (Supplementary 1). The roles of these three team members are complementary. The healthcare professional represents the clinical need and brings knowledge to the studied field. They assess the data for quality and evaluate the model's outcome for clinical usefulness. The patient or parent protects the patient's interests, is the model right for the population, and does it answer a relevant problem? Furthermore, the patient observes if the data are used correctly and in the best interest of past and future patients. Finally, the data scientist brings expertise in handling big data sets and preparing the data for analysis. Moreover, the data scientist has proficiency in different machine learning models, choosing the right model and interpreting and explaining the outcome to the team. The roles are referred to as representatives, as the project team members are seen as representatives of a larger group of colleagues with similar expertise. This notion is important because it prevents the project team members from pursuing only their interests and overseeing potential caveats and unidentified possibilities. The multidisciplinary way of working is not established just by putting a project team together. A prerequisite for the transparency of the project is the fact that

[1]Department of Neonatology, University Medical Center Utrecht, Utrecht University, UMC Utrecht, Utrecht, The Netherlands. [2]Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht University, UMC Utrecht, Utrecht, The Netherlands. ✉email: D.C.Vijlbrief@umcutrecht.nl
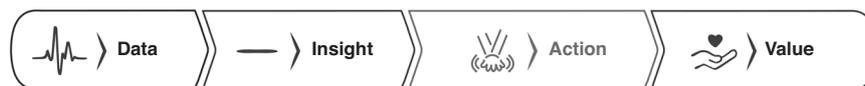
Fig. 1 Data analytics value chain. Value is created through actionable insights generated from data. This figure shows an example from cardiology, where a noted absence of a heart rhythm is acted upon by resuscitation to save a life.
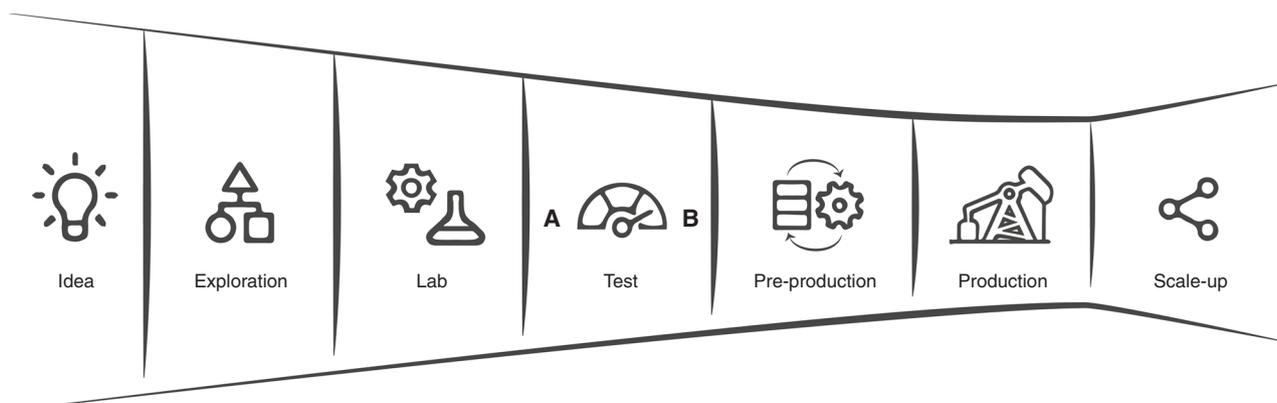
Fig. 2 Innovation funnel. Eight-phased funnel with go/no/go-gates in between. Examples of clinician involvement throughout the funnel include specification of the clinical problem (idea phase), education of data scientists on the meaning of routine care data (exploration phase), establishing the most clinically important model performance parameters (lab phase), input on clinical usability (test phase), providing input on training materials (pre-production phase), ensuring responsible use (production phase), being a clinical ambassador for the product (scale-up phase).

Table 1. An example of variation in routinely collected data, all writing variants of Augmentin® (amoxicillin and clavulanic acid) found in the electronic health records.

A	Au	Aug	Augmentin	Augementint	Augemntin
Augemtenin	Augemtin	Augentin	Augm	Augmantin	Augmebtin
Augmemtin	Augmenbtin	Augmenin	Augmenitin	Augmenitn	Augmenstin
Augmenti	Augmentien	Augmentiin	Augmentim	Augmentin\	Augmentine
Augmenting	Augmentinm	Augmentinn	Augmentint	Augmentn	Augmentu
Augmetin	Augmetnin	Augmintin	Augmmentin	Augmnentin	Augmnetin
Augnemtin	Augnmentin	Auhmentin	Aum	Aumentin	Augmentin

team members know the limits and limitations of their expertise and ask for help, including and beyond the colleagues they represent.

SEEING DATA THROUGH PATIENTS
Data are the oxygen for any data science project. Even a sampling rate of one per minute produces 10,000 samples per week of only a single continuous variable. Usually, the sampling rate is much higher. An ECG monitor can produce samples at a frequency of 500 Hz. Several studies have identified heart rate variability, the estimated difference between beats, as a marker of sepsis and mortality in a NICU.[6,7] Suppose even a single variable can hold such crucial information. Image what information is hidden in the other data. Are physiologic parameters the way the infant speaks to us, and is data science our new way of listening?

One needs to realize that patients merely "generate" data as a result of routine clinical care and used together with the clinical picture of the actual patient to provide the best care. Re-use of these data thus warrants extreme scrutiny. The task of the healthcare professional is to identify what useful data are generated by a clinical course and how to combine them (seeing data through patients).

Good quality data are essential for a successful data science project (garbage in = garbage out) and can be a real challenge while using routine care data. Even though in the EHR, the data are presented orderly, this may not be the case at the backend in the database itself (Table 1). As a result, many related tables need to be combined through different identifiers to create a valid set. Furthermore, most of the EHR consists of unstructured notes, an even more significant challenge.

A significant gap exists between what is stored and what is needed for successful data science projects, which means there is a lot of so-called pre-processing work for the data scientists before a model can be made. Pre-processing is estimated to take up to 80–90% of all the work of a data scientist. However, with the appropriate amount of time and resources, data can be sufficiently cleaned and made fit for the job. Of course, this can be done in an iterative approach, which we endorse.

SEEING PATIENTS THROUGH DATA
When useful data have been selected, re-painting the clinically relevant picture can begin. A data scientist needs to know what patient could be identified by what variables (or "features") as having a disease (yes or no) at any moment. It is, therefore, essential to come to a clear definition of the investigated disease based on available data. Choosing the right definition for the outcome variable depends on the clinical need. Sometimes a clear diagnosis is not defined and needs to be established through expert meetings.

Furthermore, identifying what part of the disease process needs data science support, early warning, diagnosis, prognosis, or outcome, is essential, as all require a different set of data. For instance, a model predicting the onset of sepsis cannot include a C-reactive protein measurement, as the blood sample holds information on the clinical suspicion.

EDUCATION AS A CATALYST FOR PROJECT PROGRESS

Becoming a neonatologist generally requires years of specialist training. Still, even for a neonatologist, defining a diagnosis can be challenging.[8,9] To develop a clinically relevant algorithm, we need the input of data scientists. However, a data scientist is trained in data, machine learning, and algorithms, not in the world of bronchopulmonary dysplasia and infant respiratory distress syndrome. The healthcare professional needs to educate the data scientist on the clinical process as well as the disease, and vice versa, as data science knowledge is usually limited for neonatologists.[10] This requires the appropriate amount of time for education. During the project, this creates a lot of positive energy and is essential for a successful project. An atmosphere where saying "I don't know" is normal and asking for help is encouraged facilitates transparency and trust. Limited knowledge, miscommunication, or misinterpretation can lead to unusable or even faulty assumptions and, therefore, algorithms.

DISCUSSION

In the UMC Utrecht, we have recently gained experience developing valuable digital health solutions using an innovation funnel approach and multidisciplinary teamwork. Entering the world of data-driven innovation and AI can overwhelm healthcare professionals. In this article, we have discussed four angles that any clinician involved in data science projects can immediately adopt, establishing a multidisciplinary way of working, seeing data through patients and patients through data, and using education as a catalyst for progress. Of course, many more challenges exist in developing and implementing digital health solutions. However, the multidisciplinarity of the project team, where every team member brings their expertise to the table, is more than capable of handling development and implementation challenges.

DATA AVAILABILITY

All relevant data can be found in the manuscript.

REFERENCES

1. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
2. Stevens, M., Wehrens, R. & de Bont, A. Epistemic virtues and data-driven dreams: on sameness and difference in the epistemic cultures of data science and psychiatry. *Soc. Sci. Med.* **258**, 113116 (2020).
3. Van De Sande, D. et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. *BMJ Health Care Inform.* **29**, e100495 (2022).
4. Haitjema, S., Prescott, T. R. & van Solinge, W. W. The Applied Data Analytics in Medicine Program: lessons learned from four years' experience with personalizing health care in an academic teaching hospital. *JMIR Form. Res.* **6**, e29333 (2022).
5. Bezemer, T. et al. A human(e) factor in clinical decision support systems. *J. Med. Internet Res.* **21**, e11732 (2019).
6. Moorman, J. R. M. D. et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. *J. Pediatr.* **159**, 900–906.e1 (2011).
7. Fairchild, K. D. et al. Septicemia mortality reduction in neonates in a heart rate characteristics monitoring trial. *Pediatr. Res.* **74**, 570–575 (2013).
8. Dong, Y. & Speer, C. P. Late-onset neonatal sepsis: recent developments. *Arch. Dis. Child. Fetal Neonatal Ed.* **100**, F257–F263 (2015).
9. Bekhof, J., Reitsma, J. B., Kok, J. H. & Van Straaten, I. H. Clinical signs to identify late-onset sepsis in preterm infants. *Eur. J. Pediatr.* **172**, 501–508 (2013).
10. Spruit, M., Dedding, T. & Vijlbrief, D. Self-service data science for healthcare professionals: a data preparation approach. *Proc. 13th Int. Jt. Conf. Biomed. Eng. Syst. Technol.* **5**, 724–734 (2020).

AUTHOR CONTRIBUTIONS

M.B., S.H., D.V., W.v.S., and J.D. contributed to the conception and design of the manuscript. All authors read and approved the final manuscript. Manuscript writing: all authors. Final approval of manuscript: all authors. Accountable for all aspects of the work: all authors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION