Statistical learning for complex data to enable precision medicine strategies

Zwep, L.B.

Chapter 2

Beyond the Randomized Clinical Trial: Innovative Data Science to Close the Pediatric Evidence Gap

Authors

Sebastiaan C. Goulooze

Laura B. Zwep

Julia E. Vogt

Elke H. J. Krekels

Thomas Hankemeier

John N. van den Anker

Catherijne A. J. Knibbe

Clinical Pharmacology & Therapeutics 2020; 107(4), 786–795

Abstract

Despite the application of advanced statistical and pharmacometric approaches to pediatric trial data, a large pediatric evidence gap still remains. Here, we discuss how to collect more data from children by using real-world data from electronic health records, mobile applications, wearables, and social media. The large datasets collected with these approaches enable and may demand the use of artificial intelligence and machine learning to allow the data to be analyzed for decision making. Applications of this approach are presented, which include the prediction of future clinical complications, medical image analysis, identification of new pediatric end points and biomarkers, the prediction of treatment nonresponders, and the prediction of placebo-responders for trial enrichment. Finally, we discuss how to bring machine learning from science to pediatric clinical practice. We conclude that advantage should be taken of the current opportunities offered by innovations in data science and machine learning to close the pediatric evidence gap.

2.1 Introduction

Historically, the evidence basis of pediatric treatments has lagged behind those in adult patients. A key aspect of this is the lack of pediatric data, which originates from the logistic, ethical, and legal challenges of performing clinical investigations in children (Brussee et al., 2016). Additionally, the pediatric population is more heterogeneous than the adult population, with maturational differences in pharmacokinetics, pharmacodynamics, and disease etiology across the pediatric age range from preterm neonates to adolescents (Brussee et al., 2016). Consequently, data collected in children within a narrow age range might still leave us with limited information regarding the treatment of children outside the studied age range. Finally, similar to other patient populations, optimal treatment will also differ for individuals within the same age group, for instance, because of obesity, genetic polymorphisms, or disease severity, and should be improved with more personalized treatment approaches (Allegaert et al., 2017).

To date, academic hospitals and industry perform clinical studies and randomized clinical trials (RCTs) on current and new drugs in children. Many academic studies focus on commonly used drugs in hospitalized patients, as the in-patient situation facilitates the collection of data. Generally, to minimize the study burden on pediatric subjects, the frequency and amount of data collection is limited and often not standardized. For example, to limit the number of venous samples, drug concentrations in plasma might be quantified in scavenged samples that were taken as part of standard of care (Krekels et al., 2017). Population pharmacometric modeling approaches have been successfully used to deal with these unbalanced data to better understand pediatric pharmacology (Brussee et al., 2016; Krekels et al., 2017). More recently, we have seen an increased use of mechanistic or physiologically based models, which leverage prior knowledge regarding the physiological changes in organ weight, blood flow,

and protein expression during a child's life (Allegaert et al., 2017; Mehrotra et al., 2016; Rostami-Hodjegan, 2012). An important aspect of such models is their improved predictive performance when used to extrapolate from adults to children (Danhof, 2015).

These pharmacometric modeling approaches are now, despite limited data, being used with success to support neonatal and pediatric drug development as well as dosing of commonly used off-label drugs (Brussee et al., 2016; Mehrotra et al., 2016; Barker et al., 2018). However, recent failures of RCTs in children have taught us that there is more to these studies than confirming model-based predictions (Momper et al., 2015). These failures have been attributed to different reasons, such as an increased placebo effect in children, different disease etiology compared with adults, and inadequate dose selection (Momper et al., 2015). Another important cause is the failure to recruit sufficient patients, which can force investigators to costly increases of the study duration or even premature termination of a study due to low feasibility of recruiting the target sample size (Joseph et al., 2015; Denhoff et al., 2015). Failed drug trials—and the general lack of pediatric clinical trials being performed particularly in primary health care—contribute to the high prevalence of off-label drug use in children, especially in the first years of life (Yackey et al., 2019). It is clear that despite the advances in approaches to data collection and analysis, a large need for additional research in pediatrics still remains.

To tackle the limitations of conventional clinical research, we need to move beyond the RCTs and their analysis with traditional statistical and advanced pharmacometric techniques. In this narrative review, we will discuss novel approaches to collecting data in pediatric patients to get more information from both clinical trials and real-world data. In addition, we will discuss how large datasets that are derived from new data collection approaches enable, and may demand, the use of innovative data science approaches, such as machine learning. Finally, we will discuss both applications and challenges to the widespread use of machine learning in pediatric medicine. Together, these innovations have the potential to greatly support our ability to generate high-quality evidence to guide optimal pediatric clinical care, thereby closing the pediatric evidence gap.

2.2 Advances in pediatric data collection

Improving our capacity for pediatric data collection is necessary for closing the pediatric evidence gap. Pediatric (randomized) clinical studies are costly and time-consuming to perform, and a sole reliance on these studies may limit our capacity for medical research in children. These studies are generally site-centric, meaning that most data is collected in a hospital or physical study site. Figure 2.1 illustrates how the capacity for pediatric data collection can be increased by moving beyond site-centric pediatric studies toward real-world data and new techniques for patient-centric data collection (Swift et al., 2018). Below we elaborate on the different opportunities and challenges (ethical and privacy) of these advances in data collection in pediatrics.

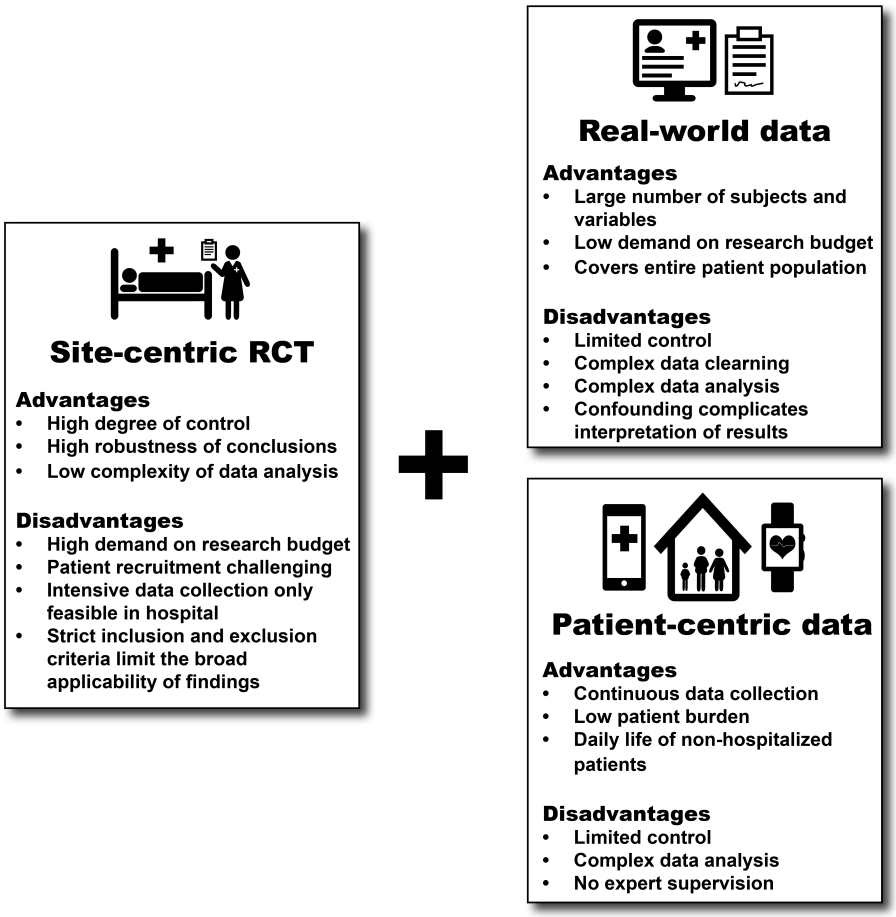

Figure 2.1: Innovative pediatric data collection beyond the site-centric randomized clinical trial (RCT). Data from these pediatric clinical studies can be supplemented by increased use of real-world pediatric data from electronic health records. Additional information can be obtained without increasing patient burden by using patient-centric data collection tools, such as mobile applications, wearables, and social media data. Site-centric RCTs refer to studies in which data collection is limited to one or more hospitals or physical study sites. Patient-centric data refers to data collected from the patient at home or during other parts of their daily routine.

2.2.1 Real-world data collection

The collection of real-world data through electronic health records (EHRs) has sharply increased in the last decade, which opens up an unprecedented potential for data collection with more subjects, more variables, and lower costs (Goldstein et al., 2016). The use of EHR data for research purposes comes with its own set of challenges, due to the large amount of data and variables to be analyzed. Machine learning techniques are often required to maximize the information extracted from EHRs. In addition to

large amounts of structured data, a part of the information in EHRs is hidden in clinical or laboratory notes, which complicates data analysis when this information is required to answer a particular question (Swift et al., 2018). To extract information from such notes into structured data, techniques like natural-language processing may provide a great opportunity for answering pediatric research questions (Nadkarni et al., 2011; Savova et al., 2016). These techniques enable analyses that would be impossible to perform on the text data itself when it would be too time-consuming to do a manual extraction of the relevant features from the text data. For example, in radiology, natural-language processing was used to automatically notate whether a certain condition or finding is mentioned within the text of the report (Pons et al., 2016). In another example, Liang and others used natural-language processing to allow the use of unstructured information from EHRs for the development of a deep-learning model for automatic pediatric diagnoses that surpassed the accuracy of junior, but not senior, physicians (H. Liang et al., 2019). Finally, the data extracted using natural-language processing might be required to identify patients eligible for inclusion in cohorts for observational research (Savova et al., 2016).

Although effectiveness research with real-world data can be problematic due to the difficulty in controlling for confounding variables and nonrandomized treatment decisions, real-world data offer many other opportunities (Swift et al., 2018; Eichler et al., 2018; Miksad et al., 2019). First, real-world data might be used to generate or select hypotheses on the most effective treatment that can then be tested in an RCT. Alternatively, real-world data might be used to confirm that the findings in a well-controlled RCT also apply to the wider, more heterogeneous pediatric population or establish that some subpopulations require additional research (Eichler et al., 2018). Additionally, real-world data can also be used to better characterize patients outside clinical studies as natural history cohorts that can subsequently be used as an external control to replace placebo arms in pediatric trials (Miksad et al., 2019). Although externally controlled studies require additional considerations to deal with potential biases compared with traditional RCTs, this approach might provide an opportunity for performing studies in cases where sufficiently powered RCTs are difficult to perform due to rarity of the indication or reluctance of parents to consent to a placebo-controlled trial (Miksad et al., 2019; Dejardin et al., 2017; Food and Drug Administration (FDA), 2019a). Finally, real-world data may be more suitable than RCTs for answering drug safety questions regarding rare adverse effects or adverse effects that present themselves years after the initial drug exposure (Eichler et al., 2018; McMahon & Pan, 2018).

To deliver the best medical practice tomorrow, it is important that we harness the full potential of the data collected today. At the moment, data in EHRs are still primarily collected for medical practice and may sometimes be ill suited for secondary use as research data. This is compounded by the fact that physicians are primarily responsible for treating patients and not for generating high-quality research data (Eichler et al., 2018). In a learning healthcare system, real-world data are not only collected to treat the individual patient but also readily usable to improve clinical practice by contributing to the generation of knowledge and innovations (Eichler et

al., 2018; C. P. Friedman et al., 2016). Examples of initiatives include the PEDSnet learning healthcare system, a large clinical data research network that currently holds data of over 6 million children from 2009 onward and has enabled the generating of real-world evidence in a variety of clinical settings, including obesity, leukemia, and long-term safety of (maternal) drug use (Forrest et al., 2014). In addition, important are initiatives like the European EHR4CR project (Moor et al., 2015) that support the integration of data from different EHR systems, as this allows the creation of larger datasets, and the external validation of findings in datasets from different sites (Goldstein et al., 2016; Eichler et al., 2018).

2.2.2 Patient-centric data collection

In addition to data from site-centric RCTs, in which most data is collected in one or more physical study sites (McMahon & Pan, 2018), the collection of patient-centric data has the potential to increase the capacity for data collection (Figure 2.1) (McMahon & Pan, 2018; C. P. Friedman et al., 2016). Patient-centric data refers to data collected from the patient at home or during other parts of their daily routine. Depending on the context, data could be collected using mobile applications, wearables, and social media. A specific advantage of patient-centric data is the increase of study data without increasing the study burden associated with additional study visits that may, in the case of children, affect their parents or caregivers as well. The opportunities of patient-centric data collection are particularly important for studying chronic diseases in children that do not require hospitalization or frequent hospital check-ups as part of their treatment. Another potential application would be the long-term follow-up of previously hospitalized patients.

Mobile applications. In its simplest form, a mobile application might be an electronic diary, designed to collect self-reported outcomes, which can be reported by children when they are beyond a certain age or by the parents in case of younger children. Compared with a paper diary, electronic diaries are reported to improve compliance with alerts and to reduce the risk of errors during data entry (Izmailova et al., 2018). In other cases, the primary aim of the application is to promote healthy behavior in the child through motivation or education, for example, in applications that help older children with self-management of asthma or type 1 diabetes (Majeed-Ariss et al., 2015). The interactions by the child and/or their parents with these applications may offer great opportunities for data collection.

Wearables. The use of wearables creates the possibility of continuous data collection in an at-home setting, which supports characterizing the intra-individual and inter-individual variability in disease and drug response, as well as quantifying exposure–response relationships for drugs in the pediatric population (Kothare et al., 2018). The latter is especially true if the clinical outcome or a surrogate end point can be quantified at home. Similar to mobile applications, the wearable itself might not only be

used to collect data but also to motivate desirable behavior. For example, Hooke et al. evaluated the use of activity trackers to promote physical activity in children with acute lymphoblastic leukemia in an effort to reduce treatment-induced fatigue (Hooke et al., 2016).

Wearables can also include biochemical sensors to noninvasively measure electrolytes, metabolites, and proteins in an at-home setting. Wearables worn on the skin can be used to measure analytes directly in sweat, but can also noninvasively extract analytes, such as proteins and glucose from the skin's interstitial fluid (Kim et al., 2019). Although many analytes of interest cannot yet be measured using wearable sensors, future developments in this area will likely expand the applicability of these techniques for patient-centric collection of pediatric biochemical data.

Although these wearables may provide great opportunities for data collection in otherwise difficult to study patient populations, like children, it is important to note that the field of clinical application of wearables is still in its infancy when we consider its clinical utility, even for adult patients (Khozin & Coravos, 2019). There are a variety of challenges that need to be met in scientific, logistic, ethical, and privacy aspects, as covered extensively by a recent review by Izmailova et al (Izmailova et al., 2018). For example, commercially available wearables frequently do not report the raw data, but only the summary or secondary data that has been processed with undisclosed and proprietary algorithms. This complicates the interpretation of wearable data, especially when collecting data from multiple types of wearables with differing terminology and data standards. For the pediatric application of wearables, additional validation will be required to ensure devices are also fit-for-purpose for children of a particular age group, and whether the data measured with these devices have the same relevance for the clinical outcome. Finally, the use of wearables by study participants might affect their behavior (e.g., they might walk more when wearing a wearable that tracks their daily step count), which could be a problem depending on the research question and design of the study. Despite these challenges, their ability for continuous data collection at low burden to the patient could provide a great opportunity in the effort to fill the pediatric evidence gap, especially if the link can be made to clinical outcomes and biomarkers.

Social media data. The use of social media has increased dramatically over the last decade. It has been reported that children who use medication might use these platforms to share experiences that are not communicated to their healthcare practitioner (Dreisbach et al., 2019). As such, social media might contain information useful to pediatric pharmacovigilance that is not available elsewhere. Recent studies explored patient reports of adverse effects on social media platforms, such as Twitter (Patel et al., 2018) and patient fora (Marshall et al., 2015). This information was explored by counting how many times different adverse effects were mentioned in combination with a certain drug. These studies could serve as a method for signal detection of rare adverse effects, or to supplement information on known adverse effects that are underestimated in children.

At the moment, the use of social media data for pharmacovigilance is still in its

infancy. In a recent study from the IMI project WEB-RADR, natural-language processing techniques that were used to automatically label social media posts with drugs and adverse effects combinations were only correct in about 40% of the cases (van Stekelenborg et al., 2019). Using these imprecise techniques, the authors found no indication that posts on general social media platforms like Facebook and Twitter would have an added value to traditional methods of pharmacovigilance. Another challenge identified in the WEB-RADR project is that some drugs are hardly discussed in social media posts, thus having little to no potential for advancing pharmacovigilance (van Stekelenborg et al., 2019). The use of social media posts in pharmacovigilance might be more beneficial with further advances in natural-language processing and by directing research efforts toward patient fora, which would carry a higher percentage of relevant posts than general social media platforms.

2.3 Ethical and Privacy Aspects of Pediatric Data Collection

Innovations in data collection will support our ability to effectively treat pediatric patients in the future, especially when the collected data is Findable, Accessible, Interoperable, Reusable to allow secondary analyses to be performed by the broader research community. These benefits need to be balanced with the right to privacy of the patients whose data are used in this research. Maintaining and further developing ethical and data security standards are crucial to ensure ongoing support by patients and their parents of data collection for research purposes (Shaw et al., 2019). Maintaining data security is particularly challenging for patient-centric data collection where sensitive data are collected on a mobile phone or wearable, as data leaks could occur when the device is lost or during data transfer from the device to the central database.

Appropriate security measures need to be in place to minimize the risk of violating the patient's privacy. In this respect, the removal of identifying information can contribute to maintaining privacy when using data for research purposes. However, when the research question requires that data from different databases are linked, some form of patient identifier might be needed to do this (Currie, 2013). A potential solution to this issue is to add a small amount of noise to the data to ensure patients cannot be identified (Currie, 2013). Another interesting approach is to "share the answers, not the data." In this case, a data analysis or model might be run on the data, and only the aggregated results are returned to the researchers.

The issue of consent is particularly complex for pediatrics. Depending on the age of the child, (written) informed consent might be obtained from the parents, the child, or both. However, in the case of reuse of the data, there are questions that remain unanswered (Taylor et al., 2017). Can the parental informed consent be considered to be valid for reuse of the data years later, even if the patient has since reached adolescence or adulthood? It is recognized that retrospectively obtaining informed consent for large datasets of observational real-world data could likely result in lengthy

and costly procedures, which would limit their use in practice (Currie, 2013). However, for some observational analyses of de-identified data, the need for informed consent can be waived by institutional review board, if appropriate privacy measures are taken (Currie, 2013).

2.4 Machine Learning for Evidence Generation

Innovations in pediatric data collections provide great opportunities for research and hold great promise in closing the pediatric evidence gap, but this promise can only be fulfilled if these data are used effectively to address clinically relevant questions. To do so is challenging due to the size and complexity of datasets collected with these novel techniques. Collecting new types of data will, therefore, go hand in hand with the increasing use of artificial intelligence and machine learning in pediatrics.

The term machine learning is often used interchangeably with the term artificial intelligence (AI). AI is an area in the discipline of computer science that aims to create intelligently perceiving, reasoning, and acting machines. A subset of AI is machine learning, which encompasses a wide range of advanced data analyses techniques. Depending on techniques used, machine learning algorithms can predict both numerical outcomes (e.g., a disease severity score) or class labels (e.g., healthy vs. diseased).

With respect to the different classes of machine learning techniques, linear models are an easy to interpret class of machine learning techniques for the analysis of structured data (Figure 2.2). Linear regression, which is the most common linear modeling technique, can be used for both prediction and hypothesis testing, but is not suitable when there are many variables in the dataset. In those cases, penalized regression techniques can be used, which have a penalty term to constrain overfitting. Examples of such techniques include lasso (Tibshirani, 1996) and ridge regression (Hoerl & Kennard, 1970). A second class of machine learning techniques are tree-based models, such as Classification and Regression Trees (Breiman et al., 1984; J. H. Friedman, 2001) and random forests (Breiman, 2001a). Depending on the specific type of technique, the output of a tree-based model might be a form of a decision tree, which can still be relatively well explained. A third class of machine learning techniques is deep learning or deep neural networks. Deep learning has been used extensively for image analysis and text mining outside the medical world and has recently started to be used on medical images and EHRs (Figure 2.2) (Miotto et al., 2017). Complex deep-learning models can have a good predictive performance when dealing with unstructured data due to flexibility of such models (Figure 2.2). However, deep-learning models are often difficult to explain, as it is generally difficult to understand how the input data leads to the model prediction.

Of note, it is important to recognize that machine learning will supplement, and not replace, traditional statistics in pediatric research. The use of traditional statistical tests or linear models might be more appropriate if the primary goal of the analysis is not to obtain a prediction model (Breiman, 2001b; Donoho, 2017). This includes situations when the goal of the analysis is hypothesis testing (“Does the treatment

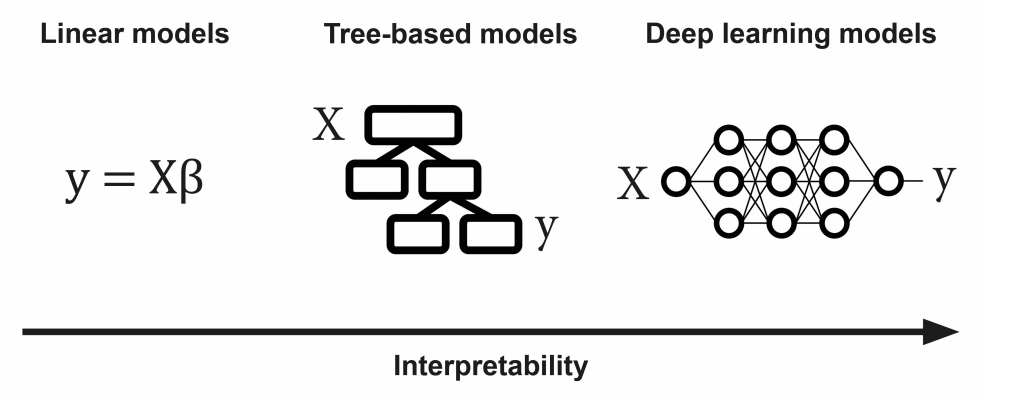

Figure 2.2: Explicability of the various machine learning techniques. On the far left, linear models have a clear explanation, but require that the data are structured. Linear regression, the most common linear modeling technique, can be used for both prediction and hypothesis testing, but is not suitable when there are many variables in the dataset. On the opposite end of the spectrum, deep-learning models are generally difficult to interpret and explain, and not suitable for hypothesis testing. However, due to the flexibility of deep-learning models, they are able to handle complex and unstructured data, such as image and text data. Depending on the data (structured or unstructured) and the goal of the analysis (raw predictive performance or testing hypotheses), different techniques will be most appropriate.

work better than placebo?") or estimation of treatment effect ("What effect does the treatment have on the outcome?"). However, there are various clinical problems in which the predictions made by machine learning can contribute to closing the pediatric evidence gap, as will be illustrated with examples in the next section.

2.5 Applications of Machine Learning in Pediatrics

The opportunities offered by the various machine learning techniques can benefit pediatric practice in a variety of ways. In this section, we will discuss different applications of machine learning in pediatrics, including: the prediction of future clinical complications, medical image analysis, identification of new pediatric end points and biomarkers, prediction of treatment nonresponders, and the prediction of placebo-responders to allow pediatric trial enrichment.

2.5.1 Predicting future clinical complications

The ability to predict clinical complications in the future can be used to deliver more personalized medicine in pediatrics. For this purpose, machine learning plays a crucial role due to its improved potential predictive performance compared with traditional statistical methods, especially when the data are unstructured or otherwise complex. Children, who are predicted to be at high risk for a certain event, can subsequently be monitored and treated more intensively. In recent research, new algorithms have been explored to make good predictions using data from previous

studies or real-world data. Box 1 shows three case studies in which machine learning techniques were used to make predictions about future clinical complications, such as childhood obesity (Dugan et al., 2015), late onset sepsis (Mani et al., 2014), and neonatal hyperbilirubinemia (Daunhawer et al., 2019).

Box 1 Prediction of clinical complications in pediatrics using machine learning

Case study 1. Childhood obesity Dugan et al. (2015) explored predictors of childhood obesity, with the aim of eventually being able to provide targeted obesity prevention for high-risk children. The answers on a dynamic questionnaire and measurements of clinical staff were mined from over 7000 children below the age of 2 years. These features were used to predict the prevalence of obesity after their second birthday. Using tree-based machine learning, an accurate model predicting childhood obesity was obtained, which included predictors like pre-existing obesity, ethnicity, height and maternal depression.

Case study 2. Neonatal sepsis Mani et al. (2014) evaluated the usefulness of different classification algorithms to predict late onset sepsis in neonates, using early results of laboratory tests and nursing observations. The best classification algorithm surpassed the clinician in both the sensitivity and specificity of predicting neonatal sepsis. After validation, clinical implementation could allow earlier treatment of sepsis while reducing the number of patients unnecessarily treated with antibiotics.

Case study 3. Neonatal hyperbilirubinemia Daunhawer et al. (2019) used machine learning techniques to predict neonatal hyperbilirubinemia. An ensemble classifier combining the logistic regression lasso and random forests was able to predict accurately whether a neonate would undergo phototherapy treatment in the next 48 hours. The predictions were made using clinical variables, such as birth weight and health information about the mother. This model could support a more personalized bilirubin monitoring approach, with more intensive monitoring of high-risk patients.

2.5.2 Medical image analysis

Deep-learning models have been particularly effective in image analysis, mainly in radiology (Yamashita et al., 2018). A deep-learning model can learn to classify images as healthy or diseased or can notate the areas in the image that correspond to organs or other anatomic structures. For example, a deep-learning model was able to identify the segmentation of white matter, gray matter, and cerebrospinal fluid in the brains of babies (Zhang et al., 2015). The automation of these tasks with a deep-learning model can reduce the time spent on an image by limiting the radiologist's task to checking and adjusting the lines drawn by the algorithm. In another example, a deep-learning model was able to identify the skeletal maturity of children by assessing hand radiographs (Larson et al., 2018). Another common application is the detection of malignant tumors in medical images, which could serve as a second opinion to detect malignancies that might have been missed by the radiologist (Suzuki et al.,

2005; M. Liang et al., 2016).

In addition to increasing efficiency, deep-learning models could also extract information from image data that is not included in the radiologist report. This would include features that are too complex and time-consuming to extract manually or features that are not currently being used in clinical decision making (Hosny et al., 2018). With automated extraction of additional information from medical images, deep learning-based image analysis can be used to perform research on imaging-based pediatric biomarkers that would not be feasible with manual image analysis.

2.5.3 Identifying end points and biomarkers in pediatrics

The development and validation of pharmacodynamic end points for children is recognized as an important methodological step in closing the evidence gap of pediatric medicine (Kelly et al., 2018). Having suitable disease-specific pharmacodynamic end points for children is essential for demonstrating efficacy and for establishing the exposure-response relationship of drugs needed for pediatric drug labeling. Additionally, these measures of patient disease severity or well-being can guide treatment decisions in clinical practice. For this, the efficacy and safety end points used in adults may not be fit-for-purpose across the pediatric age range: the clinical end point might not occur until later in life, might not be directly measurable, or the clinical presentation of the disease might differ too much from any adult counterpart (Kelly et al., 2018).

Machine learning can be used in biomarker and end-point discovery by performing variable selection and dimension-reduction when there are multiple variables considered to be potentially relevant for pediatric outcome. For example, Hartley et al. used electroencephalography data to derive a summary measure for nociceptive brain activity in infants (Hartley et al., 2017). In this example, the electroencephalography-based measure of pain was learned from the context (i.e., by comparing the response profiles after non-noxious or noxious stimulation). In another example, a supervised learning approach was used to derive a measure of iatrogenic withdrawal severity in children by combined analysis of nurse's expert opinion of the child's withdrawal severity and the observed withdrawal symptoms (Goulooze et al., 2019). Finally, machine learning may be used to identify early biomarkers that correspond to long-term clinical end points or quality of life (Bera et al., 2019). For example, a machine learning tool is currently being developed to analyze cough sound data as a digital biomarker of acute respiratory disease in children (Coravos et al., 2019).

2.5.4 Predicting treatment responders

Machine learning techniques can also be used to identify nonresponders (i.e., children who are unlikely to respond to a particular treatment). The clinical benefit lies in avoiding therapy that might give adverse effects at low chance of beneficial effects, as well as reducing the need for trial-and-error approaches for treatment personalization (Doherty et al., 2018).

For the adults, machine learning techniques have been used to predict nonresponders to drug treatment in different settings, including oncology, immunology, and postoperative pain (Doherty et al., 2018; Gram et al., 2016; Huang et al., 2018). Depending on the similarity of disease between adults and children, and the explicability and the biological plausibility of the machine learning model, models developed in adults might be also applicable in the pediatric setting after validation. In other cases, the pediatric pathophysiology might be too different or the disease might be absent in adults. In this case, efforts would be warranted to develop new machine learning models to predict drug response in children, so that they can also benefit from these innovations.

2.5.5 Predicting placebo responders to improve trial success

Prospective (randomized) clinical trials remain the gold standard to get drugs registered for the pediatric population. However, some of these RCTs fail to demonstrate efficacy in children (Momper et al., 2015). These failures have been attributed to a numbers of reasons, one of which is the high placebo response observed in indications such as depression, migraine, and bipolar disorder (Momper et al., 2015). A high placebo response would limit the ability of a trial to demonstrate efficacy or would require a very large sample size to do so. Additionally, it has been shown that younger children tend to have a stronger placebo response than older children (Weimer et al., 2013). This would make it especially difficult to demonstrate efficacy in younger children, which is problematic considering that the off-label drug use is highest in children in the first year of life (Yackey et al., 2019).

One way to limit the impact of placebo response on trial outcomes would be to identify baseline predictors of placebo response so that trials can be enriched prerandomization with subjects that are less likely to respond strongly to placebo (Momper et al., 2015). This strategy has been used in pediatric trials, resulting, for example, in the successful application for a pediatric indication of rizatriptan for acute treatment of migraine (Sun et al., 2013). For adults, it has been proposed that machine learning techniques may have better predictive power when using multiple variables to predict placebo response, as was demonstrated for depression in a geriatric population (Zilcha-Mano et al., 2018). The use of machine learning techniques to reduce the placebo response in pediatric trials might, therefore, increase the success rate of pediatric drug trials and support pediatric drug labeling.

2.6 Bringing Machine Learning to Pediatric Practice

Whereas promising, more work needs to be done before the machine learning applications mentioned in the previous section are ready for widespread clinical use in children. Methods for predicting placebo response need to be developed for different therapeutic indications and prove their worth in practice by increasing the success of pediatric registration trials (Figure 2.3, left column). Biomarkers and end points

suggested by machine learning need to be validated and supported by the relevant stakeholders (Figure 2.3, middle column). When this is the case, having better pediatric end points and biomarkers will impact not only pediatric practice, but also pediatric research. Considerable work is also required to bring a machine learning model to the clinic as a medical decision support tool, as this requires extensive external validation of the model, the development of a user-friendly software tool, and assessment of the impact of the use of this tool in clinical practice (Figure 2.3, right column). Below, we will discuss the issue of validation of machine learning models for clinical use and the particular challenges of implementing medical decision support tools in pediatric clinical practice.

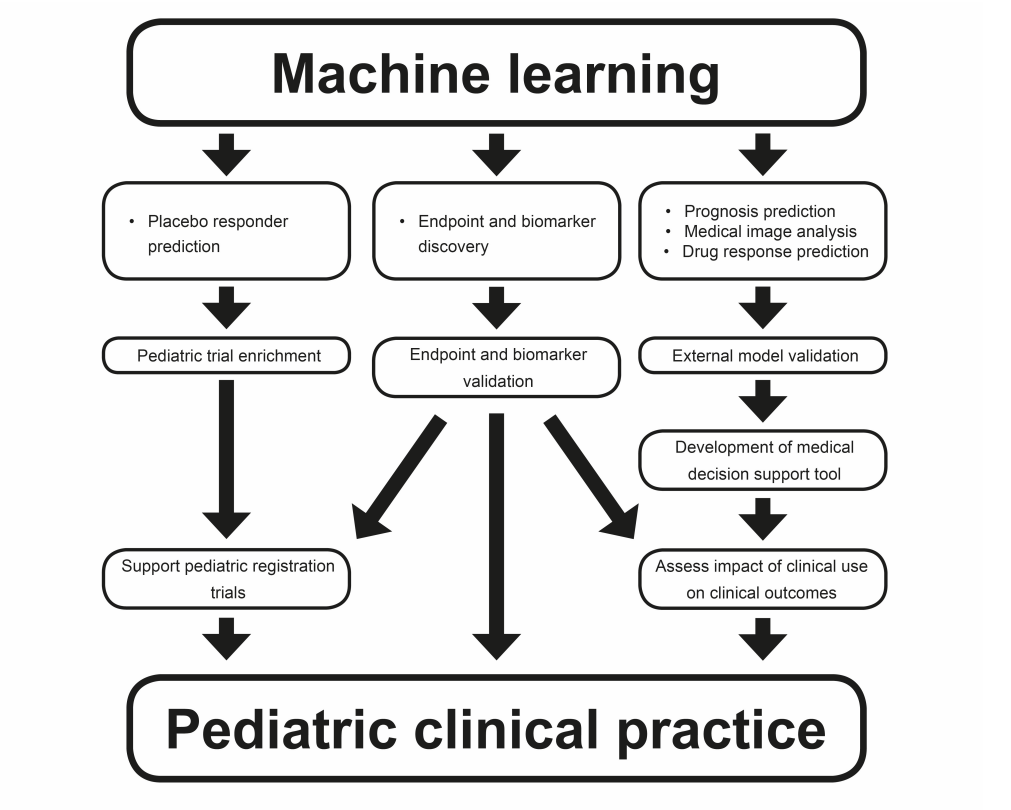

Figure 2.3: How applications of machine learning in pediatrics can support pediatric clinical practice.

2.6.1 Validation for clinical use in pediatrics

Machine learning models enable us to use complex data to achieve improved predictions of health and disease in children compared with traditional methods. However, it is important that the trained model does not "overfit" the data. An overfitted machine learning model has good predictive performance in the dataset it was trained on, but

poor performance when predicting for new cases. Validation of the model on an independent test data set is, therefore, essential to ensure the scientific quality and the clinical utility of the model (Figure 2.3, right column). Obtaining suitable datasets for this external validation can be challenging, especially in pediatric research, which underlines the importance of efforts to promote data sharing and the use of real-world data for research purposes (Ince et al., 2009).

Considering the heterogeneity of the pediatric population aged 0–18 years, it is also important to consider that a model might have a good predictive performance for children in a particular age group, but a poor performance for others (e.g., preterm neonates vs. term neonates). This risk is particularly high if certain age groups are underrepresented or absent in the dataset used to develop the model (Vayena et al., 2018). Transparency about the validity of the model and for which pediatric population this validity has been shown is, therefore, crucial.

Finally, it is important to recognize that even externally validated model predictions are not guaranteed to improve patient outcome when used in clinical practice. Some have, therefore, proposed that the clinical use of models as medical decision support tools should be supported by studies that demonstrate their impact on relevant clinical end points (Figure 2.3, right column) (Darcy et al., 2016). Considering the added difficulty to perform such trials in children, we argue that it is important to consider the need for such trials on a case-by-case basis, depending on the potential risk and benefits of the use (and nonuse) of machine learning tools in clinical decision making. In cases where dedicated pediatric trials are not feasible, modeling and simulation workflows used in pharmacometrics might be used to assess the likely clinical benefit-risk ratio of decision support tools by integrating available data from adult and pediatric patients (Bellanti et al., 2015).

2.6.2 Implementation of medical decision support tools

Implementation of findings from machine learning studies into pediatric clinical practice will not happen without focused efforts and close involvement of the various stakeholders. Currently, the widespread clinical implementation of scientific evidence is a lengthy process (> 15 years on average) and only achieved in about half of the cases (Bauer et al., 2015). Wittmeier et al. have argued in favor of systematic stepwise approaches to bring scientific knowledge to pediatric clinical practice. An important aspect of this is to engage in activities that have been shown to successfully support implementation, such as educational outreach and meetings, use of local opinion leaders, computerized reminders, audit, and feedback (Wittmeier et al., 2015). For the implementation of machine learning as a medical decision support tool in pediatrics, there are additional challenges to overcome (Figure 2.3, right column) (Shaw et al., 2019).

Because the predictions or classifications of machine learning tools can incorporate information of multiple variables, they are not as readily integrated in clinical guidelines as knowledge that relies on a single variable (e.g., age or bodyweight) for decision making. Therefore, software packages might be needed so that physicians

can easily use models in medical decision making (Figure 2.3, right column). It is important to stress that such software packages should be quick and simple to use and ideally linked to the EHR system so that there is no need for error-prone data entry of a large number of variables by the clinician.

The need to integrate machine learning tools into software packages does complicate their implementation, as they can be classified by the US Food and Drug Administration (FDA) as a medical device if the physician is not able to independently evaluate the basis of the recommendation (Food and Drug Administration (FDA), 2019b). With complex machine learning models, this is likely the case. Many software packages that provide recommendations based on models obtained with machine learning would, therefore, require lengthy regulatory approval procedures before they can be used in clinical practice.

In addition to being easy to use, the advice of the model should be explicable by the clinician. Here lies a key challenge for machine learning tools, especially for techniques like neural networks, which provide more "black box" predictions (Zorc et al., 2019). The integration of such black box predictions in clinical decision making is problematic because it means a departure from the paradigm of evidence-based medicine (Adkins, 2017). Additionally, shared decision making between the patient and physician also requires that decisions supported by machine learning tools can also be explained (Vayena et al., 2018; Zorc et al., 2019). Therefore, explicability for both the physician and the patient is likely a requirement for meaningful contributions to the decision process. Ongoing efforts to improve the explicability of complex machine learning models are, therefore, crucial to support their clinical acceptance and implementation (Cabitza et al., 2017).

2.7 Conclusions

Innovations in data collection and analysis could revolutionize many aspects of medical science and clinical practice in the upcoming decades. With the increased use of real-world data within a learning healthcare system and patient-centric data collection there is a potential to significantly expand our capacity for pediatric data collection. There are many useful potential applications of the predictive performance of machine learning models, and future work may integrate these applications with mechanistic modeling to improve understanding of the underlying biology. In addition, even though efforts are required to bring these innovations to the clinic, it is crucial that we capitalize on this opportunity to close the pediatric evidence gap.

References

Adkins, D. E. (2017, Feb). Machine learning and electronic health records: A paradigm shift. *American Journal of Psychiatry*, *174*(2), 93–94. Retrieved from https://doi.org/10.1176%2Fappi.ajp.2016.16101169 doi: 10.1176/appi.ajp.2016.16101169

Allegaert, K., Simons, S. H., Tibboel, D., Krekels, E. H., Knibbe, C. A., & van den Anker, J. N. (2017, Nov). Non-maturational covariates for dynamic systems pharmacology models in neonates, infants, and

children: Filling the gaps beyond developmental pharmacology. *European Journal of Pharmaceutical Sciences*, *109*, S27–S31. Retrieved from https://doi.org/10.1016%2Fj.ejps.2017.05.023 doi: 10.1016/j.ejps.2017.05.023

Barker, C. I. S., Standing, J. F., Kelly, L. E., Faught, L. H., Needham, A. C., Rieder, M. J., ... Offringa, M. (2018, Apr). Pharmacokinetic studies in children: recommendations for practice and research. *Archives of Disease in Childhood*, archdischild–2017–314506. Retrieved from https://doi.org/10.1136%2Farchdischild-2017-314506 doi: 10.1136/archdischild-2017-314506

Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. M. (2015, Sep). An introduction to implementation science for the non-specialist. *BMC Psychology*, *3*(1). Retrieved from https://doi.org/10.1186%2Fs40359-015-0089-9 doi: 10.1186/s40359-015-0089-9

Bellanti, F., van Wijk, R. C., Danhof, M., & Pasqua, O. D. (2015, Jul). Integration of PKPD relationships into benefit-risk analysis. *British Journal of Clinical Pharmacology*, *80*(5), 979–991. Retrieved from https://doi.org/10.1111%2Fbcp.12674 doi: 10.1111/bcp.12674

Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V., & Madabhushi, A. (2019, Aug). Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*, *16*(11), 703–715. Retrieved from https://doi.org/10.1038%2Fs41571-019-0252-y doi: 10.1038/s41571-019-0252-y

Breiman, L. (2001a, Oct 01). Random forests. *Machine Learning*, *45*(1), 5-32. Retrieved from https://doi.org/10.1023/A:1010933404324 doi: 10.1023/A:1010933404324

Breiman, L. (2001b, Aug). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3). Retrieved from https://doi.org/10.1214%2Fss%2F1009213726 doi: 10.1214/ss/1009213726

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. In (p. 368).

Brussee, J. M., Calvier, E. A. M., Krekels, E. H. J., Välitalo, P. A. J., Tibboel, D., Allegaert, K., & Knibbe, C. A. J. (2016, Jun). Children in clinical trials: towards evidence-based pediatric pharmacotherapy using pharmacokinetic-pharmacodynamic modeling. *Expert Review of Clinical Pharmacology*, *9*(9), 1235–1244. Retrieved from https://doi.org/10.1080%2F17512433.2016.1198256 doi: 10.1080/17512433.2016.1198256

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017, Aug). Unintended consequences of machine learning in medicine. *JAMA*, *318*(6), 517. Retrieved from https://doi.org/10.1001%2Fjama.2017.7797 doi: 10.1001/jama.2017.7797

Coravos, A., Khozin, S., & Mandl, K. D. (2019, Mar). Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *npj Digital Medicine*, *2*(1). Retrieved from https://doi.org/10.1038%2Fs41746-019-0090-4 doi: 10.1038/s41746-019-0090-4

Currie, J. (2013, Apr). "big data" versus "big brother": On the appropriate use of large-scale data collections in pediatrics. *Pediatrics*, *131*(Supplement_2), S127–S132. Retrieved from https://doi.org/10.1542%2Fpeds.2013-0252c doi: 10.1542/peds.2013-0252c

Danhof, M. (2015, Aug). Kinetics of drug action in disease states: towards physiology-based pharmacodynamic (PBPD) models. *Journal of Pharmacokinetics and Pharmacodynamics*, *42*(5), 447–462. Retrieved from https://doi.org/10.1007%2Fs10928-015-9437-x doi: 10.1007/s10928-015-9437-x

Darcy, A. M., Louie, A. K., & Roberts, L. W. (2016, Feb). Machine learning and the profession of medicine. *JAMA*, *315*(6), 551. Retrieved from https://doi.org/10.1001%2Fjama.2015.18421 doi: 10.1001/jama.2015.18421

Daunhawer, I., Kasser, S., Koch, G., Sieber, L., Cakal, H., Tütsch, J., ... Vogt, J. E. (2019, Mar). Enhanced early prediction of clinically relevant neonatal hyperbilirubinemia with machine learning. *Pediatric Research*, *86*(1), 122–127. Retrieved from https://doi.org/10.1038%2Fs41390-019-0384-x doi: 10.1038/s41390-019-0384-x

Dejardin, D., Delmar, P., Warne, C., Patel, K., van Rosmalen, J., & Lesaffre, E. (2017, Dec). Use of a historical control group in a noninferiority trial assessing a new antibacterial treatment: A case study and discussion of practical implementation aspects. *Pharmaceutical Statistics*, *17*(2), 169–181. Retrieved from https://doi.org/10.1002%2Fpst.1843 doi: 10.1002/pst.1843

Denhoff, E. R., Milliren, C. E., de Ferranti, S. D., Steltz, S. K., & Osganian, S. K. (2015, Oct). Factors associated with clinical research recruitment in a pediatric academic medical center—a web-based survey. *PLOS ONE*, *10*(10), e0140768. Retrieved from https://doi.org/10.1371%2Fjournal.pone.0140768 doi: 10.1371/journal.pone.0140768

Doherty, M. K., Ding, T., Koumpouras, C., Telesco, S. E., Monast, C., Das, A., ... Schloss, P. D. (2018, May). Fecal microbiota signatures are associated with response to ustekinumab therapy among crohn's disease patients. *mBio*, *9*(2). Retrieved from https://doi.org/10.1128%2Fmbio.02120-17 doi: 10.1128/mbio.02120-17

Donoho, D. (2017, Oct). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766. Retrieved from https://doi.org/10.1080%2F10618600.2017.1384734 doi: 10.1080/10618600.2017.1384734

Dreisbach, C., Koleck, T. A., Bourne, P. E., & Bakken, S. (2019, May). A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *International Journal of Medical Informatics*, *125*, 37–46. Retrieved from https://doi.org/10.1016%2Fj.ijmedinf.2019.02.008 doi: 10.1016/j.ijmedinf.2019.02.008

Dugan, T. M., Mukhopadhyay, S., Carroll, A., & Downs, S. (2015). Machine Learning Techniques for Prediction of Early Childhood Obesity. *Applied Clinical Informatics*, *06*(03), 506–520. doi: 10.4338/aci-2015-03-ra-0036

Eichler, H.-G., Bloechl-Daum, B., Broich, K., Kyrle, P. A., Oderkirk, J., Rasi, G., ... Paris, V. (2018, Oct). Data rich, information poor: Can we use electronic health records to create a learning healthcare system for pharmaceuticals? *Clinical Pharmacology & Therapeutics*, *105*(4), 912–922. Retrieved from https://doi.org/10.1002%2Fcpt.1226 doi: 10.1002/cpt.1226

Food and Drug Administration (FDA). (2019a). *Clinical Decision Support Software. Draft Guidance for Industry and Food and Drug Administration Staff* (Tech. Rep.). Silver Spring, MD: Food and Drug Administration. Retrieved from http://www.fda.gov/downloads/MedicalDevices/.../UCM263366.pdf

Food and Drug Administration (FDA). (2019b). *Rare Diseases: Natural History Studies for Drug Development* (Tech. Rep. No. March 2019). Silver Spring, MD: Food and Drug Administration. Retrieved from https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

Forrest, C. B., Margolis, P. A., Bailey, L. C., Marsolo, K., Beccaro, M. A. D., Finkelstein, J. A., ... Kahn, M. G. (2014, Jul). PEDSnet: a national pediatric learning health system. *Journal of the American Medical Informatics Association*, *21*(4), 602–606. Retrieved from https://doi.org/10.1136%2Famiajnl-2014-002743 doi: 10.1136/amiajnl-2014-002743

Friedman, C. P., Allee, N. J., Delaney, B. C., Flynn, A. J., Silverstein, J. C., Sullivan, K., & Young, K. A. (2016, Nov). The science of learning health systems: Foundations for a new journal. *Learning Health Systems*, *1*(1), e10020. Retrieved from https://doi.org/10.1002%2Flrh2.10020 doi: 10.1002/lrh2.10020

Friedman, J. H. (2001, Oct). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. Retrieved from https://projecteuclid.org/journals/annals-of-statistics/volume-29/issue-5/Greedy-function-approximation-A-gradient-boosting-machine/10.1214/aos/1013203451.full doi: 10.1214/aos/1013203451

Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2016, May). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, *24*(1), 198–208. Retrieved from https://doi.org/10.1093%2Fjamia%2Focw042 doi: 10.1093/jamia/ocw042

Goulooze, S. C., Ista, E., van Dijk, M., Hankemeier, T., Tibboel, D., Knibbe, C. A., & Krekels, E. H. (2019, Oct). Supervised multidimensional item response theory modeling of pediatric iatrogenic withdrawal symptoms. *CPT: Pharmacometrics & Systems Pharmacology*, *8*(12), 904–912. Retrieved from https://doi.org/10.1002%2Fpsp4.12469 doi: 10.1002/psp4.12469

Gram, M., Erlenwein, J., Petzke, F., Falla, D., Przemeck, M., Emons, M., ... Drewes, A. (2016, Jul). Prediction of postoperative opioid analgesia using clinical-experimental parameters and electroencephalography. *European Journal of Pain*, *21*(2), 264–277. Retrieved from https://doi.org/10.1002%2Fejp.921 doi: 10.1002/ejp.921

Hartley, C., Duff, E. P., Green, G., Mellado, G. S., Worley, A., Rogers, R., & Slater, R. (2017, May). Nociceptive brain activity as a measure of analgesic efficacy in infants. *Science Translational Medicine*, *9*(388). Retrieved from https://doi.org/10.1126%2Fscitranslmed.aah6122 doi: 10.1126/scitranslmed.aah6122

Hoerl, A. E., & Kennard, R. W. (1970, Feb). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. Retrieved from https://doi.org/10.1080%2F00401706.1970.10488634 doi: 10.1080/00401706.1970.10488634

Hooke, M. C., Gilchrist, L., Tanner, L., Hart, N., & Withycombe, J. S. (2016, Jan). Use of a fitness tracker to promote physical activity in children with acute lymphoblastic leukemia. *Pediatric Blood & Cancer*, *63*(4), 684–689. Retrieved from https://doi.org/10.1002%2Fpbc.25860 doi: 10.1002/pbc.25860

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018, May). Artificial intelligence in radiology. *Nature Reviews Cancer*, *18*(8), 500–510. Retrieved from https://doi.org/10.1038%2Fs41568-018-0016-5 doi: 10.1038/s41568-018-0016-5

Huang, C., Clayton, E. A., Matyunina, L. V., McDonald, L. D., Benigno, B. B., Vannberg, F., & McDonald, J. F. (2018, Nov). Machine learning predicts individual cancer patient responses to therapeutic drugs

with high accuracy. *Scientific Reports*, *8*(1). Retrieved from https://doi.org/10.1038%2Fs41598-018-34753-5 doi: 10.1038/s41598-018-34753-5

Ince, I., de Wildt, S. N., Tibboel, D., Danhof, M., & Knibbe, C. A. (2009, Mar). Tailor-made drug treatment for children. *Drug Discovery Today*, *14*(5-6), 316–320. Retrieved from https://doi.org/10.1016%2Fj.drudis.2008.11.004 doi: 10.1016/j.drudis.2008.11.004

Izmailova, E. S., Wagner, J. A., & Perakslis, E. D. (2018, Apr). Wearable devices in clinical trials: Hype and hypothesis. *Clinical Pharmacology & Therapeutics*, *104*(1), 42–52. Retrieved from https://doi.org/10.1002%2Fcpt.966 doi: 10.1002/cpt.966

Joseph, P. D., Craig, J. C., & Caldwell, P. H. (2015, Feb). Clinical trials in children. *British Journal of Clinical Pharmacology*, *79*(3), 357–369. Retrieved from https://doi.org/10.1111%2Fbcp.12305 doi: 10.1111/bcp.12305

Kelly, L. E., Sinha, Y., Barker, C. I. S., Standing, J. F., & Offringa, M. (2018, Apr). Useful pharmacodynamic endpoints in children: selection, measurement, and next steps. *Pediatric Research*, *83*(6), 1095–1103. Retrieved from https://doi.org/10.1038%2Fpr.2018.38 doi: 10.1038/pr.2018.38

Khozin, S., & Coravos, A. (2019, Apr). Decentralized trials in the age of real-world evidence and inclusivity in clinical investigations. *Clinical Pharmacology & Therapeutics*, *106*(1), 25–27. Retrieved from https://doi.org/10.1002%2Fcpt.1441 doi: 10.1002/cpt.1441

Kim, J., Campbell, A. S., de Ávila, B. E.-F., & Wang, J. (2019, Feb). Wearable biosensors for healthcare monitoring. *Nature Biotechnology*, *37*(4), 389–406. Retrieved from https://doi.org/10.1038%2Fs41587-019-0045-y doi: 10.1038/s41587-019-0045-y

Kothare, P. A., Jadhav, P. R., Gupta, P., Harrelson, J. C., & Dickmann, L. (2018, Jun). Harnessing the potential of emerging digital health and biological sampling technologies for clinical drug development: Promise to reality. *Clinical Pharmacology & Therapeutics*, *104*(6), 1125–1135. Retrieved from https://doi.org/10.1002%2Fcpt.1100 doi: 10.1002/cpt.1100

Krekels, E. H., van Hasselt, J. C., van den Anker, J. N., Allegaert, K., Tibboel, D., & Knibbe, C. A. (2017, Nov). Evidence-based drug treatment for special patient populations through model-based approaches. *European Journal of Pharmaceutical Sciences*, *109*, S22–S26. Retrieved from https://doi.org/10.1016%2Fj.ejps.2017.05.022 doi: 10.1016/j.ejps.2017.05.022

Larson, D. B., Chen, M. C., Lungren, M. P., Halabi, S. S., Stence, N. V., & Langlotz, C. P. (2018, Apr). Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs. *Radiology*, *287*(1), 313–322. Retrieved from https://doi.org/10.1148%2Fradiol.2017170236 doi: 10.1148/radiol.2017170236

Liang, H., Tsui, B. Y., Ni, H., Valentim, C. C. S., Baxter, S. L., Liu, G., ... Xia, H. (2019, Feb). Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nature Medicine*, *25*(3), 433–438. Retrieved from https://doi.org/10.1038%2Fs41591-018-0335-9 doi: 10.1038/s41591-018-0335-9

Liang, M., Tang, W., Xu, D. M., Jirapatnakul, A. C., Reeves, A. P., Henschke, C. I., & Yankelevitz, D. (2016, Oct). Low-dose CT screening for lung cancer: Computer-aided detection of missed lung cancers. *Radiology*, *281*(1), 279–288. Retrieved from https://doi.org/10.1148%2Fradiol.2016150063 doi: 10.1148/radiol.2016150063

Majeed-Ariss, R., Baildam, E., Campbell, M., Chieng, A., Fallon, D., Hall, A., ... Swallow, V. (2015, Dec). Apps and adolescents: A systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. *Journal of Medical Internet Research*, *17*(12), e287. Retrieved from https://doi.org/10.2196%2Fjmir.5043 doi: 10.2196/jmir.5043

Mani, S., Ozdas, A., Aliferis, C., Varol, H. A., Chen, Q., Carnevale, R., ... Weitkamp, J.-H. (2014, Mar). Medical decision support using machine learning for early detection of late-onset neonatal sepsis. *Journal of the American Medical Informatics Association*, *21*(2), 326–336. Retrieved from https://doi.org/10.1136%2Famiajnl-2013-001854 doi: 10.1136/amiajnl-2013-001854

Marshall, S. A., Yang, C. C., Ping, Q., Zhao, M., Avis, N. E., & Ip, E. H. (2015, Oct). Symptom clusters in women with breast cancer: an analysis of data from social media and a research study. *Quality of Life Research*, *25*(3), 547–557. Retrieved from https://doi.org/10.1007%2Fs11136-015-1156-7 doi: 10.1007/s11136-015-1156-7

McMahon, A. W., & Pan, G. D. (2018, Jun). Assessing drug safety in children — the role of real-world data. *New England Journal of Medicine*, *378*(23), 2155–2157. Retrieved from https://doi.org/10.1056%2Fnejmp1802197 doi: 10.1056/nejmp1802197

Mehrotra, N., Bhattaram, A., Earp, J. C., Florian, J., Krudys, K., Lee, J. E., ... Sinha, V. (2016, Apr). Role of quantitative clinical pharmacology in pediatric approval and labeling. *Drug Metabolism and Disposition*, *44*(7), 924–933. Retrieved from https://doi.org/10.1124%2Fdmd.116.069559 doi: 10.1124/dmd.116.069559

Miksad, R. A., Samant, M. K., Sarkar, S., & Abernethy, A. P. (2019, May). Small but mighty: The use of real-world evidence to inform precision medicine. *Clinical Pharmacology & Therapeutics*, *106*(1), 87–90. Retrieved from https://doi.org/10.1002%2Fcpt.1466 doi: 10.1002/cpt.1466

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2017, May). Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, *19*(6), 1236–1246. Retrieved from https://doi.org/10.1093%2Fbib%2Fbbx044 doi: 10.1093/bib/bbx044

Momper, J., Mulugeta, Y., & Burckart, G. (2015, Jun). Failed pediatric drug development trials. *Clinical Pharmacology & Therapeutics*, *98*(3), 245–251. Retrieved from https://doi.org/10.1002%2Fcpt.142 doi: 10.1002/cpt.142

Moor, G. D., Sundgren, M., Kalra, D., Schmidt, A., Dugas, M., Claerhout, B., ... Coorevits, P. (2015, Feb). Using electronic health records for clinical research: The case of the EHR4cr project. *Journal of Biomedical Informatics*, *53*, 162–173. Retrieved from https://doi.org/10.1016%2Fj.jbi.2014.10.006 doi: 10.1016/j.jbi.2014.10.006

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011, Sep). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544–551. Retrieved from https://doi.org/10.1136%2Famiajnl-2011-000464 doi: 10.1136/amiajnl-2011-000464

Patel, R., Belousov, M., Jani, M., Dasgupta, N., Winokur, C., Nenadic, G., & Dixon, W. G. (2018, Feb). Frequent discussion of insomnia and weight gain with glucocorticoid therapy: an analysis of twitter posts. *npj Digital Medicine*, *1*(1). Retrieved from https://doi.org/10.1038%2Fs41746-017-0007-z doi: 10.1038/s41746-017-0007-z

Pons, E., Braun, L. M. M., Hunink, M. G. M., & Kors, J. A. (2016, May). Natural language processing in radiology: A systematic review. *Radiology*, *279*(2), 329–343. Retrieved from https://doi.org/10.1148%2Fradiol.16142770 doi: 10.1148/radiol.16142770

Rostami-Hodjegan, A. (2012, May). Physiologically based pharmacokinetics joined with in vitro–in vivo extrapolation of ADME: A marriage under the arch of systems pharmacology. *Clinical Pharmacology & Therapeutics*, *92*(1), 50–61. Retrieved from https://doi.org/10.1038%2Fclpt.2012.65 doi: 10.1038/clpt.2012.65

Savova, G., Pestian, J., Connolly, B., Miller, T., Ni, Y., & Dexheimer, J. W. (2016). Natural language processing: Applications in pediatric research. In *Translational bioinformatics* (pp. 231–250). Springer Singapore. Retrieved from https://doi.org/10.1007%2F978-981-10-1104-7_12 doi: 10.1007/978-981-10-1104-7_12

Shaw, J., Rudzicz, F., Jamieson, T., & Goldfarb, A. (2019, Jul). Artificial intelligence and the implementation challenge. *Journal of Medical Internet Research*, *21*(7), e13659. Retrieved from https://doi.org/10.2196%2F13659 doi: 10.2196/13659

Sun, H., Bastings, E., Temeck, J., Smith, P. B., Men, A., Tandon, V., ... Rodriguez, W. (2013, Mar). Migraine therapeutics in adolescents. *JAMA Pediatrics*, *167*(3), 243. Retrieved from https://doi.org/10.1001%2Fjamapediatrics.2013.872 doi: 10.1001/jamapediatrics.2013.872

Suzuki, K., Li, F., Sone, S., & Doi, K. (2005, Sep). Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network. *IEEE Transactions on Medical Imaging*, *24*(9), 1138–1150. Retrieved from https://doi.org/10.1109%2Ftmi.2005.852048 doi: 10.1109/tmi.2005.852048

Swift, B., Jain, L., White, C., Chandrasekaran, V., Bhandari, A., Hughes, D. A., & Jadhav, P. R. (2018, May). Innovation at the intersection of clinical trials and real-world data science to advance patient care. *Clinical and Translational Science*, *11*(5), 450–460. Retrieved from https://doi.org/10.1111%2Fcts.12559 doi: 10.1111/cts.12559

Taylor, M. J., Dove, E. S., Laurie, G., & Townend, D. (2017, Nov). When can the child speak for herself? the limits of parental consent in data protection law for health research. *Medical Law Review*, *26*(3), 369–391. Retrieved from https://doi.org/10.1093%2Fmedlaw%2Ffwx052 doi: 10.1093/medlaw/fwx052

Tibshirani, R. (1996, Jan). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. Retrieved from https://doi.org/10.1111%2Fj.2517-6161.1996.tb02080.x doi: 10.1111/j.2517-6161.1996.tb02080.x

van Stekelenborg, J., Ellenius, J., Maskell, S., Bergvall, T., Caster, O., Dasgupta, N., ... Pirmohamed, M. (2019, Aug). Recommendations for the use of social media in pharmacovigilance: Lessons from IMI WEB-RADR. *Drug Safety*, *42*(12), 1393–1407. Retrieved from https://doi.org/10.1007%2Fs40264-019-00858-7 doi: 10.1007/s40264-019-00858-7

Vayena, E., Blasimme, A., & Cohen, I. G. (2018, Nov). Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine*, *15*(11), e1002689. Retrieved from https://doi.org/10.1371%2Fjournal.pmed.1002689 doi: 10.1371/journal.pmed.1002689

Weimer, K., Gulewitsch, M. D., Schlarb, A. A., Schwille-Kiuntke, J., Klosterhalfen, S., & Enck, P. (2013, Apr). Placebo effects in children: a review. *Pediatric Research*, *74*(1), 96–102. Retrieved from https://

doi.org/10.1038%2Fpr.2013.66 doi: 10.1038/pr.2013.66
Wittmeier, K. D. M., Klassen, T. P., & Sibley, K. M. (2015, Apr). Implementation science in pediatric health care. *JAMA Pediatrics*, *169*(4), 307. Retrieved from https://doi.org/10.1001%2Fjamapediatrics.2015.8 doi: 10.1001/jamapediatrics.2015.8
Yackey, K., Stukus, K., Cohen, D., Kline, D., Zhao, S., & Stanley, R. (2019, Mar). Off-label medication prescribing patterns in pediatrics: An update. *Hospital Pediatrics*, *9*(3), 186–193. Retrieved from https://doi.org/10.1542%2Fhpeds.2018-0168 doi: 10.1542/hpeds.2018-0168
Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018, Jun). Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, *9*(4), 611–629. Retrieved from https://doi.org/10.1007%2Fs13244-018-0639-9 doi: 10.1007/s13244-018-0639-9
Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., & Shen, D. (2015, Mar). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, *108*, 214–224. Retrieved from https://doi.org/10.1016%2Fj.neuroimage.2014.12.061 doi: 10.1016/j.neuroimage.2014.12.061
Zilcha-Mano, S., Roose, S. P., Brown, P. J., & Rutherford, B. R. (2018, Jun). A machine learning approach to identifying placebo responders in late-life depression trials. *The American Journal of Geriatric Psychiatry*, *26*(6), 669–677. Retrieved from https://doi.org/10.1016%2Fj.jagp.2018.01.001 doi: 10.1016/j.jagp.2018.01.001
Zorc, J. J., Chamberlain, J. M., & Bajaj, L. (2019, Jul). Machine learning at the clinical bedside—the ghost in the machine. *JAMA Pediatrics*, *173*(7), 622. Retrieved from https://doi.org/10.1001%2Fjamapediatrics.2019.1075 doi: 10.1001/jamapediatrics.2019.1075

Acknowledgement

The authors thank Margot van Ark for assisting in the preparation of the figures.