Chapter 15
Health Systems Management

Check for updates

Adam B. Wilcox and Bethene D. Britt

After reading this chapter, you should know the answers to these questions:

- How has artificial intelligence (AI) been used in health systems management? What are the primary areas where it has been applied? What have been the benefits of AI for health systems?
- What factors should be considered when applying AI algorithms for health systems management? How can the choice of data affect their use?
- What are the challenges in matching the cognitive tasks of interpreting data with the capabilities of AI algorithms? How can the modeling of data and the framing of the decision task affect the ability to discover insights, either through human cognition or AI application?
- How can governance be applied effectively in the adoption and application of AI algorithms?

Promise of AI in Health Systems

Health care is both complex and information-intensive. Multiple disciplines are combined to care for patients, with each having a specific role with specific information needed to support that role. Coordination across roles also involves active

A. B. Wilcox (✉)
Washington University in St. Louis School of Medicine, St. Louis, MO, USA
e-mail: a.wilcox@wustl.edu

B. D. Britt
UW Medicine, University of Washington, Seattle, WA, USA
e-mail: betbritt@uw.edu

use of information, for assessment, interpretation, comparison and communication. Care must be coordinated both for the individual patient but also for the healthcare delivery system, which increases the complexity substantially as it must consider the interaction of multiple individual patients. This complexity is reflected in various tasks where healthcare professionals must coordinate multiple sources of information and make decisions or predictions based on this information. For example, a nurse manager needs to consider the expected complexity of cases and patient volume for a unit when determining how many nurses need to be scheduled for a shift. Managers of operating rooms also need to make estimations of case complexity and time when considering scheduling for surgery to optimize resources. Clinics may over-schedule appointments based on a predicted or estimated no-show rate to optimize the total number of patients seen. These are all examples of prediction activities in the management of health systems that may benefit from the use of AI methods to quickly organize and interpret information, and are in addition to how AI can be used to make interpretations for recommended care with an individual patient.

There have been specific examples of successful adoption of various methods of AI for health systems management, ranging from resource allocation, demand and use prediction, and scheduling optimization. A review of clinical literature identifies the following published examples, which are representative of many more unpublished experiences of others in applying AI for health systems management:

- Aktas et al. reported on the development of a decision support system to improve resource allocation to be applied in a radiology department [1]. Using a **Bayesian belief network** to represent conditional dependencies, they were able to analyze the relationships of key variables affecting system efficiency for resource allocation.
- Gartner and Padman developed a **Naive Bayes classifier** to assist in early diagnosis group determination that can improve resource allocation decisions [2].
- Dennis et al. demonstrated that **artificial neural networks** were effective at predicting trauma volume and acuity within emergency departments, based on seasonality-related measures such as time of year, day of week, temperature and precipitation levels [3]. These results were consistent with heuristics related to trauma prediction, but were more precise in predicting "trauma seasons" across multiple centers.
- Lee et al. used information collected from electronic health records for emergency department patients to predict disposition decisions with logistic regression and machine learning algorithms [4]. These predictions can be used to reduce boarding delays by prompting the initiation of admission processes where needed.
- McCoy et al. demonstrated that **time-series machine learning** methods could be applied to date, census and discharge data to predict daily discharge volumes [5]. Prediction errors outside 1 standard deviation occurred only about 5% of days,

indicating the AI methods could help in predicting volumes and matching resources to volumes.

- Vermeulen et al. developed adaptive rules for scheduling computer tomography scanning that adjusts to multiple patient and system features [6]. These rules showed an ability to better adjust to variations in volume according to different resource levels.
- Kontio et al. used machine learning algorithms to predict patient acuity levels that could support resource allocation [7]. They extracted information directly from demographic data, admission information, and concepts from radiology and pathology notes from previous days to classify patients with heart problems in five different acuity categories for the current day.

In some areas of health systems management, enough examples have been published to perform systematic reviews to identify both the breadth and patterns of success:

- Bellini et al. performed a systematic review of AI applications in operating room optimization [8]. They identified successful studies in predicting procedure duration; in coordinating post-acute care unit availability; and in cancellation prediction. All of these areas can improve surgery scheduling efficiency.

However, with all of these examples and studies, widespread, consistent and successful adoption of AI in health systems management remains elusive. A review of different risk prediction models performed by Wehbe et al. noted their modest performance, infrequent use, lack of evidence for improvement and barriers to implementation. They recommended that improved approaches to machine learning adoption be determined to overcome performance barriers and allow local customization of rules, which are a recognized challenge [9].

Understanding the cognitive aspects of applied AI in health systems management may provide a path forward for more successful adoption. There are numerous examples demonstrating that successful use of AI for improvement in these areas is possible; the challenge now is to identify the features that are prerequisites for success, and to demonstrate that the success can be reproduced.

One way to achieve this goal is to improve the understanding of the actual process of AI application in health systems management. Effective application of information in health systems practice is complex, giving rise to a whole field of clinical and health informatics focused on improving the use of information in the area. As that field moved from demonstration of success to broader adoption, a deeper understanding of the interrelationships among information sources, technology, and human users has demanded different methods focused on deeper understanding of interactions [10, 11]. These lessons from clinical informatics, and specifically in decision support (which arguably includes the application of AI to health systems management), become increasingly important. The balance of this chapter therefore focuses on providing that deeper understanding through a description of two applications of information to health systems management. We also include a discussion

of governance approaches that can be used to ensure guided application of AI capabilities; this also matches the pattern of recommendations for adoption for clinical decision support by Wright et al. [12].

The first example relates to predicting no-shows for outpatient scheduling, with qualitative insights gathered from discussions with one of the authors who was a developer of the system. The second example relates to an analysis of testing approaches for monitoring devices and outcomes in heart failure patients.

Example: Outpatient Scheduling

The first example relates to an algorithm developed for predicting whether patients would arrive at clinics for scheduled appointments, in order to help staff appropriately schedule other patients to improve operational efficiency. This work was done at Massachusetts General Hospital in Boston, as described by Patrick Cronin who was working on process improvement with the Department of Medical Dermatology [13]. A critical management issue for the department was a patient rate for missed appointments that had reached 20%, which is significant and in line with general estimates nationally [14]. This meant that in a daily schedule that may include 100 appointments, 20 of those patients would not arrive as scheduled. The patients would either not arrive for the appointment (~12% "no-show" rate), or would cancel the appointment on the same day (same-day cancellation). This created important problems for the division because each no-show represented a point in the schedule with only costs but no revenue. While some same-day cancellation appointments could be filled with same-day appointment scheduling, the no-show appointments were unanticipated and therefore could not be refilled with other patients at the last minute.

Institutions have tried various methods to address the challenge of missed appointments and no-shows. One approach has been to remind patients of their appointment at least a day in advance, which can both prompt patients to attend or identify cancellations in advance. However, this approach generally requires additional staff to implement by contacting each patient beforehand, and studies have been mixed in showing differences due to reminders [14]. Another approach is to schedule additional overlapping appointments in the place of no-show appointments if they can be anticipated. This has obvious challenges if the prediction is wrong, however. If the patient does show up for the appointment, the clinic staff would then need to adapt the schedule for the rest of the day or until another gap in the schedule occurs to fit in the extra patient.

The approach applied in this example ("Smart Booking") was to create a prediction algorithm to predict patients who were most likely to miss appointments, and allow schedulers to create overlapping appointments for that day to fill the schedule. The prediction first involved identifying factors that may be related to no-shows and

were available as data for predictions. These could be based on factors about the patient (e.g., age; insurance type) or the appointment itself (appointment type; days between appointment scheduling and appointment). Also available from the longitudinal database was a 15-year history of appointment scheduling data. This allowed a patient's history of no-shows and arrivals also to be included in the prediction model, and these variables were shown to be the most significant predictors.

The result was a Smart Booking algorithm that provided a daily recommendation for available overbookings to schedulers that were used to increase appointment bookings. It was initially implemented as a randomized controlled trial (randomized by clinic session) that measured both the number of arrived patients per clinic session and the perception by physicians of how busy the clinical session was (measured by a simple scale between "too slow" and "too busy"). The results were that the Smart Booking algorithm increased the number of patients booked for appointments by an average of 0.5 arrived patients per session, with physician perception of busyness being similar between the sessions with Smart Booking (and thus more patients) and without.

The success of this project led to other related AI-enabled interventions that varied in approach, clinical domain, outcome measures, and prediction goal. These included targeted reminder phone calls or patient navigation to decrease no-shows in other clinical departments [15–17], as well as prediction rule development for hospital readmissions [18, 19].

Discussions with the author of the main study have offered important lessons learned and insights that are important in understanding the overall reproducibility and sustainability of the approach. First, the data were seen as far more important than the algorithm that was used. That is, there was a greater impact in using the prior history of visits and the appointment types than in using more complicated algorithms like neural networks instead of logistic regression. The simplest algorithm performed similarly to more complicated AI algorithms, and had the benefit of being explainable to the clinic staff who interacted with the intervention (see Chap. 8). The actual implementation was more simple than robust, with the main algorithm running on a single desktop computer of the developer, at least for the studied version of the implementation. This was in part possible due to an internal development approach leveraging the electronic health record (EHR) at the institution at the time, which allowed for a more flexible but less sustainable approach. With the later implementation of a commercial EHR at the institution, the developed algorithm was removed. The commercial EHR provides its own predictive models that could be applied, but there have been challenges in the product's implementation and outcomes when studied [20]. In this study, the customized components of the AI implementation (choice of predictor data, flexibility in implementation) were critical to success, but this approach made it difficult to both replicate and sustain. This suggests an important factor explaining why consistent and successful adoption of AI in health systems management is challenging, because the maintenance effort to keep AI effective is often expensive.

Example: Device Monitoring

The second example is not an actual application of AI in health systems management, but rather a deep analysis of data related to outcomes for a high-risk patient population. However, the understanding of how these data were analyzed is important to recognize how data and cognitive aspects of data and information must be appropriately considered in applications of AI. This analysis was performed by one of the authors (AW) and Dr. Claudius Mahr at the University of Washington Medical Center in Seattle, WA.

In this example, a discordance in laboratory tests for monitoring **intravenous unfractionated heparin (IV-UFH)** was identified by clinicians performing a quality improvement review of a series of cases with negative outcomes [21, 22]. The laboratory tests were measuring activated **partial thromboplastin time (PTT)** and **anti-factor Xa (anti-Xa)**, both of which by measuring IV-UFH can help guide anticoagulation therapy to prevent bleeding and clotting complications. This is critical for patients who have received mechanical circulatory support, such as **left-ventricular assist device (LVAD)** implantation.

For individuals with end stage heart failure, assistive technologies are needed to improve cardiac function. A common device used is a left ventricular assist device (LVAD), which is a battery-operated mechanical pump that is surgically placed in the heart, in the left ventricle or main pumping chamber. The LVAD helps the left ventricle pump blood to the rest of the body. It is often used either temporarily until a patient receives a heart transplant, or as a long-term treatment for heart failure.

Because the device is surgically implanted in the circulatory system, it can lead to clotting of the blood which can then lead to severe outcomes such as strokes or myocardial infarctions when the blood clot moves into arteries and restricts blood flow to vital organs. To prevent this, patients are given medicine such as heparin, which helps prevent the blood from clotting. However, too much heparin can lead to other complications such as internal bleeding. As a result, it is important to balance the heparin dosing and amount of heparin in the blood at a level where both clots and internal bleeding are avoided. Maintaining this balance is challenging, and involves regular laboratory measures.

A study at one institution of 200 patients over 8 years with end stage heart failure who had received LVADs showed the challenges in managing treatment [23]. The survival rate for all patients was below 50% after 4 years, indicating the severe health problems of those receiving the treatment. The most common cause of death was stroke, with 32% of deaths. Three of the five most common adverse events were internal bleeding, stroke, and re-exploration for bleeding. This indicates the importance of managing clotting and bleeding among LVAD patients, as the most common causes of death and adverse events were related to it.

PTT is a laboratory test of blood that measures the time it takes for a blood clot to form. A common therapeutic range for PTT is 60–100 s for clotting. Anti-Xa is a test that measures the amount of heparin in the blood by measuring its inhibition of factor Xa activity, which is part of the clotting process. A common therapeutic range

for anti-Xa is 0.3–0.7 IU/mL. Historically, PTT has been the more common test for clotting risk, but anti-Xa has become used increasingly. It is expected that PTT and anti-Xa would be concordant, or that when one shows a high clotting risk by a value above the therapeutic range, the other would as well. However, this is not always the case, and when there is a difference in measure it is important to know clinically which to respond to [21].

In a review of a series of patients with complications after LVAD implantation, clinicians at UW were concerned that this lack of concordance between PTT and anti-Xa could be important in patient management, and wanted an analysis of test values along with bleeding events among patients over a period of time. Over 6500 paired PTT and anti-Xa measures were identified among a patient population, and matched to bleeding events that occurred. Initially we looked at concordance measures directly as a matrix, within and above the therapeutic range for each test (Fig. 15.1). We also plotted the individual tests in a scatter plot to identify broader trends (Fig. 15.2).

To identify whether either PTT or anti-Xa were able to separate the bleeding and non-bleeding population, we plotted frequency curves for each of the measures, but

Fig. 15.1 Concordance matrix for PTT and anti-Xa measures. The blue central percentage listed in each cell gives the percent of all tests in the matrix in that cell. For example, the bottom left cell indicates that 10% of all paired measures had both an Anti Xa value of <0.3 and PTT values of 0.6. The upper left number in each cell represents the total tests in the cell, with the non-bleed count in the upper right, and the bleed count and rate at the bottom. If the tests were concordant, you would expect to see most of the test in the lower left, middle, or upper right cells, indicating the tests are low, within range, or high at the same time. In this matrix, however, the largest number of tests in the middle right cell indicates a discordance with anti-Xa in the therapeutic range (0.3–0.7 IU/mL) and PTT values above the therapeutic range (>100 s)
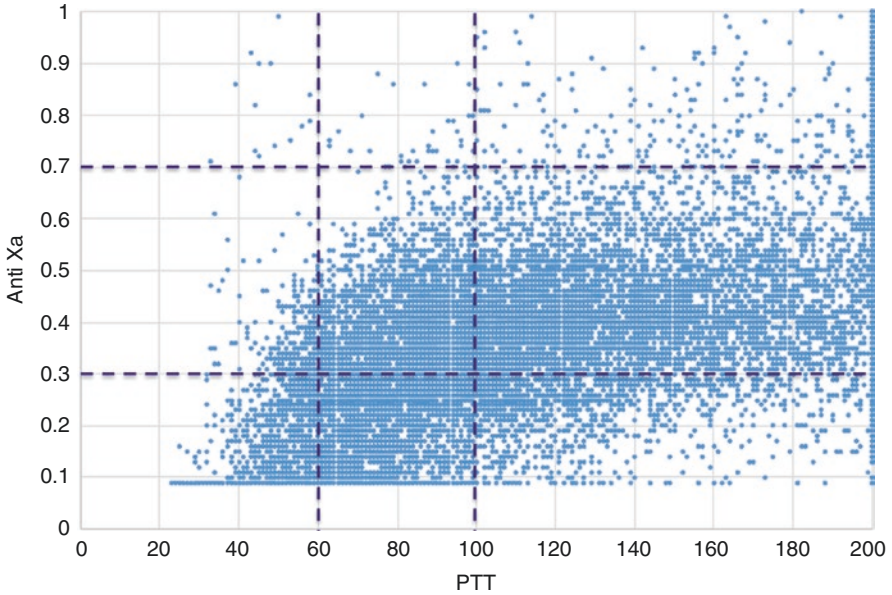
Fig. 15.2 Discordance scatter plot, which shows overall trends of PTT values being skewed above the therapeutic range, and anti-Xa values skewing below the therapeutic range. Lab tests for PTT values above 200 s had values of ">200" and were plotted as 201; anti-Xa values less than 0.1 IU/mL had values of "<0.1" and were plotted as 0.09
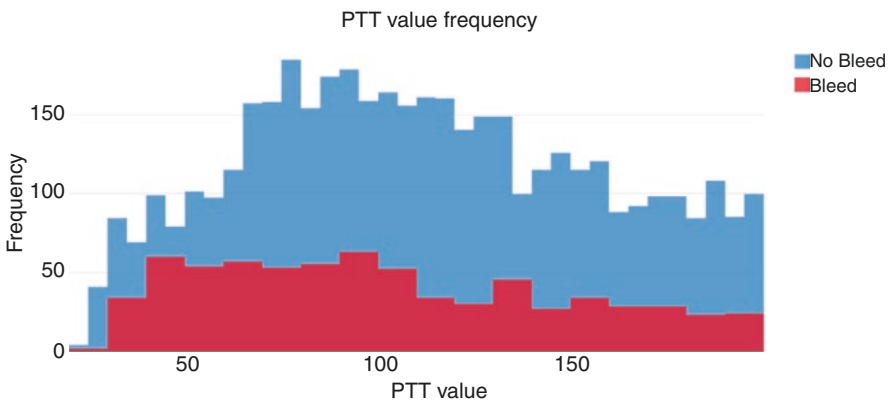

Fig. 15.3 PTT value frequency. There was no clear difference among the frequency curves that would show PTT as discriminating between bleeding risks

these showed nothing significant (Figs. 15.3 and 15.4). We also plotted PTT and anti-Xa values identifying bleeding events and calculated regression lines between the two populations (Fig. 15.5); while this showed a measurable difference, it was not convincing enough to warrant a different approach in monitoring.
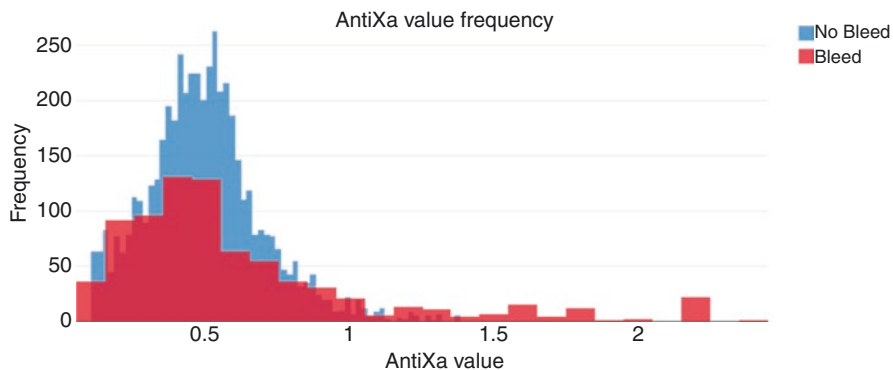
Fig. 15.4 Anti-Xa value frequency. Similar to PTT values, there was no clear discriminating difference for bleeding risks among anti-Xa values
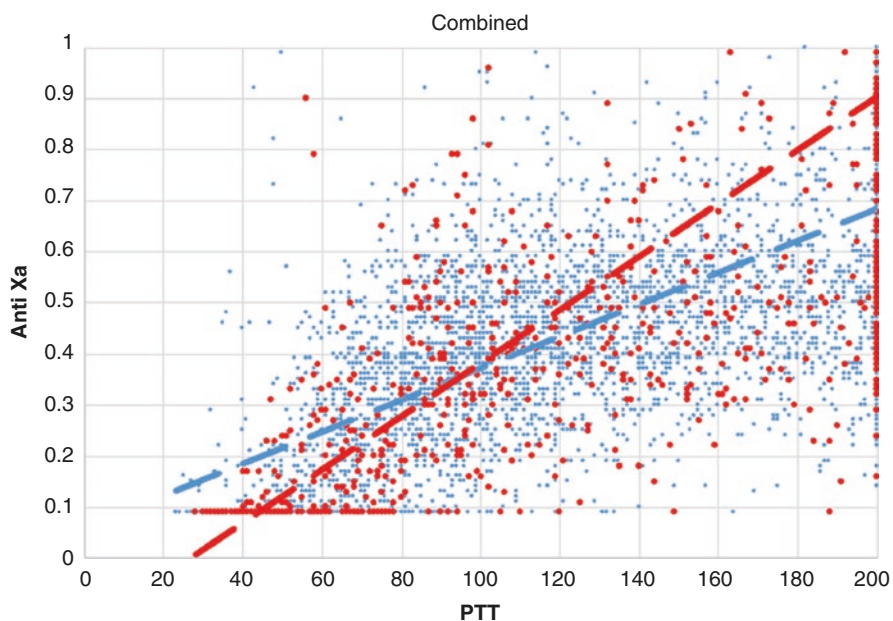

Fig. 15.5 Scatter plot with regression lines for bleeding and non-bleeding events among PTT and anti-Xa measures. The regression lines showed a measurable but unconvincing difference

Finally, we modeled the values in terms of *how they would be used in clinical decision-making*, rather than individually as values. That is, a PTT or anti-Xa result is used in making a decision to adjust therapy when it is above a threshold value that indicates a bleeding risk and need for the adjustment. To model this, we considered each value along the range of values the measure as a potential threshold point, and calculated the *bleed rate* as the proportion of bleeding events for measures at or
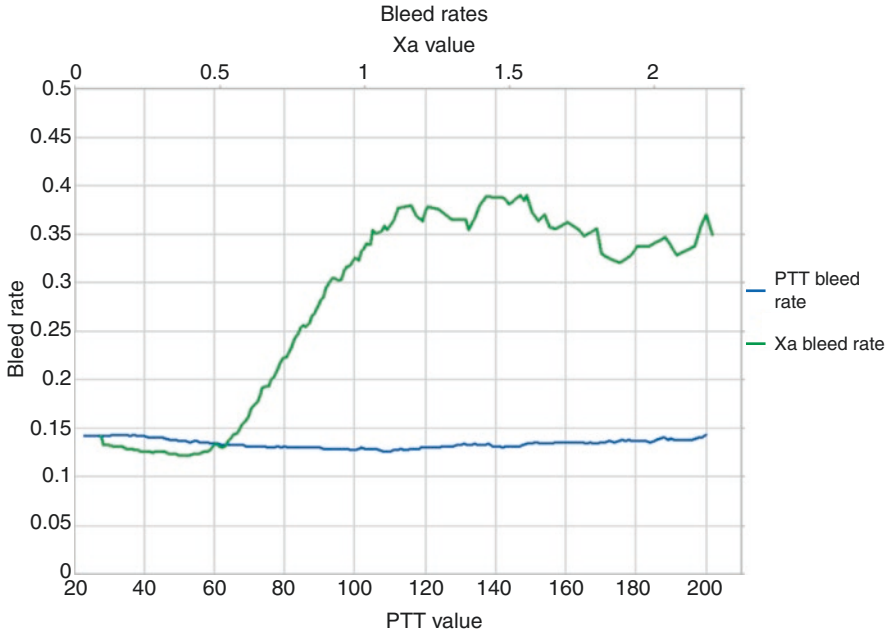
Fig. 15.6 Bleed rates calculated at varying threshold values. This approach modeled the clinical decisions by calculating the bleed rate above a threshold value, and then computing this rate for each PTT and anti-Xa value. The PTT threshold values were not helpful in discriminating bleed risks, while the anti-Xa values showed a differentiation as the values moved from the therapeutic range to high. At the highest values, there were fewer and fewer events to calculate the bleeding rate, and the rates became more variable

above that value. Modeling this threshold determination rate showed effectively how the measures would perform at different threshold value choices. As shown in Fig. 15.6, there was clear separation in the performance of anti-Xa in comparison to PTT for predicting bleeding as the threshold moved from the higher therapeutic range to the high ranges (around 0.7 for anti-Xa and 100 for PTT). Identifying this difference was important for determining the clinical approach in using anti-Xa for monitoring bleeding risk rather than PTT at the institution.

While this example did not directly use AI in creating a model, it is instructive regarding the challenges of using AI models for three primary reasons. First, there was a clear difference in the two tests and how they could be used clinically, but this wasn't apparent from many of the standard comparison approaches that are often used to evaluate AI algorithms. Rule-based approaches would be similar to a concordance matrix in their calculation approach, and regression or statistical models would be similar to the frequency distributions or regression lines. It took careful consideration of the clinical decision-making process to identify the trend. This could be identified by AI algorithms but likely only if the variable of the threshold bleed rate were determined a priori.

The second challenge relates to the actual data. These data were in many ways simpler than data that may often be used in EHR mining, in that they were numeric laboratory values. However, it took expert understanding to realize that the non-numeric values in the database of ">200" for PTT and "<0.1" for anti-Xa should be included. We did not perform a separate analysis that excluded these values, so we cannot for certain say that they needed to be included, but they contained meaning as if they were numeric values, specifically that they were above or below the therapeutic range. If nothing else, decision rules that may use the numeric values should be able to interpret how the non-numeric comparatives should be applied in the rules.

The third challenge is more subtle but also more important. This relates to how the analytic solution was reached. In the end, the analysis did not reveal something as a new hypothesis, that there was a difference in the PTT and anti-Xa values that could be important for predicting bleeding. The data analysis only verified a pattern that had been seen when experts reviewed data from a series of cases with adverse events. The pattern identification from experts was much more efficient than the complicated data analysis that verified the pattern. The verification was important, but the initial hypothesis drawn from a smaller set of examples was more important in identifying the hypothesis in the first place. Often such hypotheses are developed after observing a trend in the data that seems to indicate a significant issue. For example, if the adverse event rate was near 50% (high but not extreme for LVADs), the probability of seeing 8 patients with adverse events in a row would be 1/256 (2^{-8}), or <0.4%. This is similar to having 8 coin flips all end up "heads", which is remarkable. However, when considered among many events as a run of 8, the probability is much higher. In terms of coin flips, a run of 8 heads in a row among 200 coin flips is not nearly as rare, with a not-so-remarkable 32% probability of occurring at least once. This suggests that similar random sequences that appear rare may lead experts to perform a deeper analysis of an issue in a way that identifies underlying patterns that were not seen before. In comparing cognitive capabilities of human beings to AI methods, people may be better at identifying patterns in complex data, and worse at interpreting sequences that occur amid data reflecting random variation.

These two examples highlight important challenges when using AI in health systems management, which may be indicated by the challenges of demonstrating consistent and successful adoption of AI approaches. The second example demonstrates how the underlying data or hypothesis development can be challenging for AI development. The first example demonstrates how even when successfully implemented, AI solutions can require significant additional management and monitoring to be sustained. Other examples in AI show similar challenges [24]. Often the focus on AI is about algorithm performance, which is understandable since the algorithm is what makes it work. However, once the algorithms have been demonstrated as correct in output (which many algorithms in health systems management have), the primary challenge becomes implementation and sustainable maintenance. As a result, **governance** of resources to manage AI should be a major focus. This migration of focus from performance to governance is similar for rule-based clinical decision support (CDS) as mentioned above. Governance is critical for managing AI solutions in health systems.

Governance

Implementing AI in health systems requires practical interaction with the many systems necessary to get any technology in production. There are three main areas of governance that intersect in order to manage an AI model effectively: *Corporate Governance*, *IT Governance* and *Data Governance*.

- *Corporate Governance* provides the overall clinical and operational strategy and ethical framework for implementing AI models.
- *IT Governance* provides for the prioritization of technology investments, including the work necessary for predictive models, as well as for clinical or operational ownership and stewardship of the models.
- *Data Governance* provides governance over the underlying data that is critical for the initial development and ongoing management of the predictive models.

All three areas are necessary and must be integrated in order for AI models to be successfully able to meet the overall intended goals of the health system. (NOTE: Both IT and Data Governance are considered to be elements of Corporate Governance. We'll discuss each in turn below.)

Corporate Governance   In our current environment, it is crucial that organizations develop and communicate a point of view regarding the ethical use of AI models. Fortunately, health care has a lot of experience with ethical risk management, and the Harvard Business Review offered up this long history as a model for other industries as they consider the application of AI models.

> Leaders should take inspiration from health care, an industry that has been systematically focused on ethical risk mitigation since at least the 1970s. Key concerns about what constitutes privacy, self-determination, and informed consent, for example, have been explored deeply by medical ethicists, health care practitioners, regulators, and lawyers [25].

The opportunity here is to take advantage of (hopefully) existing corporate structures and committees within a health system that are already engaged with ethical decision-making, and to ensure that AI models are brought under this same governance. These structures need to weigh in and provide overall direction and guardrails that the other two areas of governance can follow. For example, clinical committees that already consider decision support systems, including ways to minimize alert fatigue, should be engaged in discussions about clinical AI models. Similarly, committees currently addressing issues of access to care and equity should be engaged in discussions and provide direction on how to evaluate models for bias.

IT Governance   IT Governance is a framework for making decisions about what technology investments will be made, how those technologies will be managed, and ultimately to ensure the delivery of value to the organization. AI models should be brought under IT governance and control and identified key stakeholders should be able to establish conditions for the ethical and appropriate use of models that meet

the Corporate Governance directives. At the same time, health systems are being pressured to be agile from innovators and researchers who are discovering new opportunities to improve patient care and delivery.

Organizations like Mayo Clinic ("Mayo") and UW Medicine ("UWM") have been working on governance models for AI. Mayo Clinic recently shared their governance model that includes integrating the EHR AI governance with an existing Clinical Decision Support Subcommittee. They described the role of this committee as overseeing the overall implementation and priorities, but specify that "Specialty Practice Committee approval needed for both phases and will be governing body for ongoing use/maintenance." The development phases are described below:

Phase I:
- Configure model, migrate and activate in production
- Run in background several months to localize/train
- Validate to check performance against expected results

Phase II:
- Operationalize by building tools (e.g., BPAs, dashboards)
- Develop workflows, communicate and train practice
- Maintain, annually evaluate performance and relevance

This approach has many parallels with the governance model developed at UW Medicine. A new subcommittee was formed with clinical and technical membership, named the Predictive Analytics Oversight Committee. This committee is integrated into the IT Governance committee structure, but is responsible for oversight of the overall AI implementation process and strategy. Each model has a requirement that it be "clinically/operationally led and managed," and the appropriate committee and leadership must be engaged as sponsors. The phases for the work have a similar path, outlined below (Fig. 15.7):
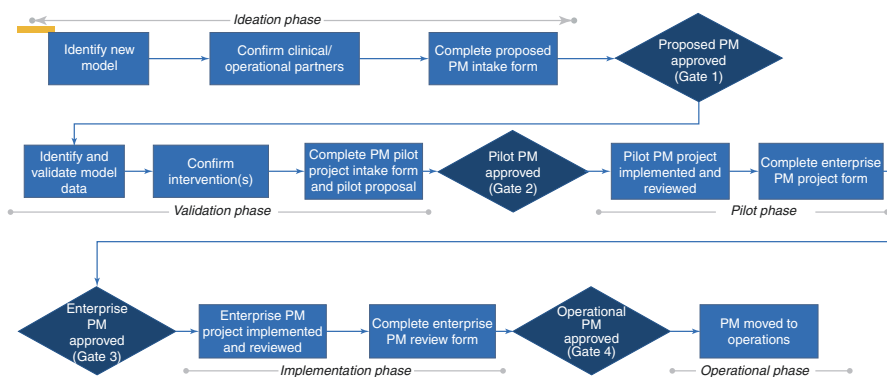

Fig. 15.7 Predictive analytics review phases. This pathway was created to assist governance of an organization's predictive analytics implementation products. *PM* predictive model

Models are independently evaluated at multiple steps in this process, including specific evaluations for patient outcomes, bias and impact to provider experience. Given the limited resources to implement models, the evaluation also factors into prioritization and schedules for implementing a specific model. Similar to the Mayo approach, the Operational Phase requires ongoing regular evaluation of the models to confirm the model is still relevant and performant.

Data Governance Most current AI approaches have been around for decades but it took large datasets and fast processors for them to find the levels of adoption seen recently (see Chap. 2). An early tenet of predictive analytics was that large data sets could overcome issues with data quality, but there are recent examples where that has been disproven [20]. AI models need high quality and timely data, and healthcare data requires appropriate data security. Healthcare data suffers from many data quality, currency and security issues:

- data collection by EHR's has been focused on the data necessary for billing purposes
- data are mostly collected and captured from busy human beings under stress
- much of the truly valuable data is captured in unstructured notes and resistant to easy feature engineering
- healthcare systems are still investing in the level of automated data entry and integration that other industries take for granted
- many organizations have defaulted to data silos in order to address data security requirements

The Data Management Book of Knowledge (DMBOK) describes Data Governance as "the exercise of authority and control over the management of data assets" [26]. A successful predictive model program in health care is dependent on maturing data governance and management to ensure optimal quality data, that is available to the right audience, with the appropriate performance and security controls.

Data pipelines will need to be architected, developed and managed in order to meet the timing needs of AI models. Critical data used by predictive models needs to be identified and real time data quality controls and alerts may be needed to monitor and manage these data over time. This monitoring can serve as an early warning that a predictive model's performance will degrade. Finally, the development of predictive models can be scaled and accelerated with the widespread adoption of meta-data management technologies, such as data dictionaries, data lineage, report catalogs, data quality dashboards, etc. This self-service infrastructure for using data assets enables a much larger range of users to identify and leverage data to develop new AI models in their own domain areas. The development of **citizen data scientists** is a significant enablement function that can have a high reward.

Operations and Maintenance As stated above, the final step in the AI implementation process is a formal transition to an operational state. The step prior to moving a model to an operational state calls for a final evaluation of the model, and

the completion of an operational intake form. This step collects information about the dependent data fields, the clinical/operational owners, an agreed upon cadence for regularly reviewing and updating the model, and ensures that the model and its descriptive information is added to a published portfolio. The dependent data fields are identified as critical fields within the data governance process so that they can be specifically reviewed for changes, and the impact of broader changes that affect data can be appropriately managed in the affected models.

While the application of governance to AI in health systems management is complex due to the complexity of the algorithms and the data, much of the governance is similar to clinical decision support governance, which has been more established. Wright et al. reviewed governance structures across multiple institutions in managing clinical decision support applications [12]. They identified six recommended practices for CDS governance, which while named differently, map closely to the governance stages described here. These practices include considering and monitoring impact, assessing changes, and ongoing monitoring. Due to the deep data dependency of AI models, the monitoring must extend more deeply into the data rather than just changes to the rules as recommended for CDS, but otherwise the practices are similar.

Concluding Remarks

In this chapter, we have discussed the use of AI or predictive analytics in health systems management. We identified studies showing that they can be successfully implemented, but contrasted that with reviews that showed challenges in broader adoption. To elucidate these challenges, we gave two detailed examples that explained the challenges both in terms of understanding data complexity and implementation. We have argued the importance of governance due to these challenges, and documented a verified governance approach for implementing and maintaining AI applications. Our primary conclusion is that successful use of intelligent systems in health systems management has been demonstrated and is possible, but it requires appropriate management of its use. As institutions were applying clinical decision support applications that had been demonstrated at individual organizations, Kawamoto et al. studied factors related to the success of CDS implementations [27]. They found that factors related to implementation were significant in determining whether they were successful or not. As demonstrated by these examples and governance recommendations, the success factors are similar; it is less about what is being implemented, and more about how it is applied.

Questions for Discussion

- What is an example of an application of AI in health systems management? What is the beneficial effect of that application on the health system?
- How are the terms "artificial intelligence" and "predictive analytics" similar? What are some differences in meaning between them?

- What are some of the challenges that may be faced when implementing a predictive analytics solution for resource management in a health system? How might these challenges be different from implementing a standard rule-based clinical decision support application?
- What are the different phases of governance for AI implementation? For each phase, what might be the consequences if the phase is not used correctly?

Further Reading

Wehbe RM, Khan SS, Shah SJ, Ahmad FS. Predicting high-risk patients and high-risk outcomes in heart failure. Heart Fail Clin. 2020;16(4):387–407.

- This paper reviews different applications of AI to health systems management and identifies some of the challenges. These challenges are important to be addressed by governance.

Bellini V, Guzzon M, Bigliardi B, Mordonini M, Filippelli S, Bignami E. Artificial Intelligence: a new tool in operating room management. Role of machine learning models in operating room optimization. J Med Syst. 2019;44(1):20.

- This paper reviews different applications of AI in operating room optimization, which is a common area for AI application in health systems management. The review covers various areas of the operating room workflow where AI has been applied.

Wright A, Sittig DF, Ash JS, Bates DW, Feblowitz J, Fraser G, Maviglia SM, McMullen C, Nichol WP, Pang JE, Starmer J, Middleton B. Governance for clinical decision support: case studies and recommended practices from leading institutions. J Am Med Inform Assoc. 2011;18(2):187–94.

- This review of decision support implementations gives specific recommendations for governance, which are similar to the governance recommendations here.

Blackman R. A Practical guide to building ethical AI. Harv Bus Rev. 2020; https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai.

- This article describes steps that companies, including those in healthcare, can take to ensure AI is implemented in a way that addresses ethical risks of using AI.

References

1. Aktas E, Ulengin F, Önsel Ş. A decision support system to improve the efficiency of resource allocation in healthcare management. Socio Econ Plan Sci. 2007;41:130–46.
2. Gartner D, Padman R. Improving hospital-wide early resource allocation through machine learning. Stud Health Technol Inform. 2015;216:315–9.
3. Dennis BM, Stonko DP, Callcut RA, Sidwell RA, Stassen NA, Cohen MJ, et al. Artificial neural networks can predict trauma volume and acuity regardless of center size and geography: a multicenter study. J Trauma Acute Care Surg. 2019;87(1):181–7.

4. Lee S-Y, Chinnam RB, Dalkiran E, Krupp S, Nauss M. Prediction of emergency department patient disposition decision for proactive resource allocation for admission. Health Care Manag Sci. 2020;23(3):339–59.

5. McCoy TH, Pellegrini AM, Perlis RH. Assessment of time-series machine learning methods for forecasting hospital discharge volume. JAMA Netw Open. 2018;1(7):e184087.

6. Vermeulen IB, Bohte SM, Elkhuizen SG, Lameris H, Bakker PJM, La Poutré H. Adaptive resource allocation for efficient patient scheduling. Artif Intell Med. 2009;46(1):67–80.

7. Kontio E, Airola A, Pahikkala T, Lundgren-Laine H, Junttila K, Korvenranta H, et al. Predicting patient acuity from electronic patient records. J Biomed Inform. 2014;51:35–40.

8. Bellini V, Guzzon M, Bigliardi B, Mordonini M, Filippelli S, Bignami E. Artificial intelligence: a new tool in operating room management. role of machine learning models in operating room optimization. J Med Syst. 2019;44(1):20.

9. Wehbe RM, Khan SS, Shah SJ, Ahmad FS. Predicting high-risk patients and high-risk outcomes in heart failure. Heart Fail Clin. 2020;16(4):387–407.

10. Kaplan B. Evaluating informatics applications--some alternative approaches: theory, social interactionism, and call for methodological pluralism. Int J Med Inform. 2001;64(1):39–56.

11. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform. 2008;41(2):387–92.

12. Wright A, Sittig DF, Ash JS, Bates DW, Feblowitz J, Fraser G, et al. Governance for clinical decision support: case studies and recommended practices from leading institutions. J Am Med Inform Assoc JAMIA. 2011;18(2):187–94.

13. Cronin PR, Kimball AB. Success of automated algorithmic scheduling in an outpatient setting. Am J Manag Care. 2014;20(7):570–6.

14. Kheirkhah P, Feng Q, Travis LM, Tavakoli-Tabasi S, Sharafkhaneh A. Prevalence, predictors and economic consequences of no-shows. BMC Health Serv Res. 2016;16:13.

15. Shah SJ, Cronin P, Hong CS, Hwang AS, Ashburner JM, Bearnot BI, et al. Targeted reminder phone calls to patients at high risk of no-show for primary care appointment: a randomized trial. J Gen Intern Med. 2016;31(12):1460–6.

16. Percac-Lima S, Cronin PR, Ryan DP, Chabner BA, Daly EA, Kimball AB. Patient navigation based on predictive modeling decreases no-show rates in cancer care. Cancer. 2015;121(10):1662–70.

17. Hwang AS, Atlas SJ, Cronin P, Ashburner JM, Shah SJ, He W, et al. Appointment "no-shows" are an independent predictor of subsequent quality of care and resource utilization outcomes. J Gen Intern Med. 2015;30(10):1426–33.

18. Cronin PR, Greenwald JL, Crevensten GC, Chueh HC, Zai AH. Development and implementation of a real-time 30-day readmission predictive model. AMIA Annu Symp Proc AMIA Symp. 2014;2014:424–31.

19. Greenwald JL, Cronin PR, Carballo V, Danaei G, Choy G. A novel model for predicting rehospitalization risk incorporating physical function, cognitive status, and psychosocial support using natural language processing. Med Care. 2017;55(3):261–6.

20. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med. 2021;181(8):1065–70.

21. Saifee N, Mahr C, Garcia D, Estergreen J, Sabath D. PTT and anti-Xa activity in adult mechanical circulatory support patients at a large academic medical center. Am J Clin Pathol. 2018;149(suppl_1):S174–5.

22. Adatya S, Uriel N, Yarmohammadi H, Holley CT, Feng A, Roy SS, et al. Anti-factor Xa and activated partial thromboplastin time measurements for heparin monitoring in mechanical circulatory support. JACC Heart Fail. 2015;3(4):314–22.

23. Tsiouris A, Paone G, Nemeh HW, Borgi J, Williams CT, Lanfear DE, et al. Short and long term outcomes of 200 patients supported by continuous-flow left ventricular assist devices. World J Cardiol. 2015;7(11):792–800.

24. STAT. Once billed as a revolution in medicine, IBM's Watson Health is sold off in parts. STAT. 2022. https://www.statnews.com/2022/01/21/ibm-watson-health-sale-equity/. Accessed 1 Feb 2022.
25. Blackman R. A practical guide to building ethical AI. Harv Bus Rev. 2020. https://hbr.org/2020/10/a-practical-guide-to-building-ethical-ai.
26. International D. DAMA-DMBOK: data management body of knowledge. 2nd ed. Basking Ridge, NJ: Technics Publications; 2017. 590 p.
27. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765.