Perspective

Prognostic models will be victims of their own success, unless...

Matthew C. Lenert[1], Michael E. Matheny[2,3,4,5], and Colin G. Walsh[6,7,8]

[1]Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, USA, [2]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [3]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [4]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [5]Geriatric Research Education and Clinical Care, Tennessee Valley Health System, Department of Veterans Affairs, Nashville, Tennessee, USA, [6]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [7]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA, and [8]Department of Psychiatry, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Corresponding Author: Matthew C. Lenert, MS, Department of Biomedical Informatics, Vanderbilt University, 2525 West End Ave, Suite 1475, Nashville, TN 37203, USA; matthew.c.lenert@vanderbilt.edu

Received 16 April 2019; Revised 8 July 2019; Editorial Decision 19 July 2019; Accepted 22 July 2019

ABSTRACT

Predictive analytics have begun to change the workflows of healthcare by giving insight into our future health. Deploying prognostic models into clinical workflows should change behavior and motivate interventions that affect outcomes. As users respond to model predictions, downstream characteristics of the data, including the distribution of the outcome, may change. The ever-changing nature of healthcare necessitates maintenance of prognostic models to ensure their longevity. The more effective a model and intervention(s) are at improving outcomes, the faster a model will appear to degrade. Improving outcomes can disrupt the association between the model's predictors and the outcome. Model refitting may not always be the most effective response to these challenges. These problems will need to be mitigated by systematically incorporating interventions into prognostic models and by maintaining robust performance surveillance of models in clinical use. Holistically modeling the outcome and intervention(s) can lead to resilience to future compromises in performance.

Key words: predictive modeling, model updating, learning health system

Artificial intelligence and machine learning change how patients and healthcare providers interact.[1–3] Many hope that predictive models are one of the technologies that improve patient health by giving insight to our current or future health, thereby promoting patient-provider-caregiver decisions that improve clinical outcomes and quality of life.[4–7] Predictive models show great promise in a number of applications, such as reducing time to accurate diagnosis, reducing complications through automated monitoring and prognosis, and empowering preventive medicine through personalized forecasting.[8–15]

Predictive models have a life cycle. This life cycle is iterative and nonlinear. In the simplest form of the model life cycle, it begins with model development, updating, or validation. Initial model development has a large and well-established literature.[16–21] Next, model developers and information technology professionals implement their models into the workflow of patients, clinicians, or administrators, so that the models might be acted on. The model implementation and deployment phase requires a different set of skills from model development. This phase is also one that investigators have studied in the informatics and implementation science literature, although this field is less mature than model development.[22–28] Within this stage, explicit consideration for how and which downstream actions could be triggered by the model implementation are often neglected.[29,30] Once the model is implemented, and impacted pro-

cess workflows stabilize, the model enters the surveillance and maintenance phase.[31–35] This phase requires skills from both the development and the implementation phase.

Surveillance and maintenance are key to preventing degradation of model performance over time.[32,36] Models degrade in healthcare, because those models attempt to characterize biological, business, or behavioral processes with unknown or unmeasured confounders.[37] Those models also rely on readily available data that cannot fully describe the underlying patient case mix, practice of medicine, and work environment. Known deterministic formulae (eg, force = mass × acceleration) rarely govern the processes informatics practitioners attempt to model, leaving those models vulnerable to changes in the processes they attempt to describe.[1] Impactful changes in the data process can also manifest as changes in documentation requirements, medical practice, or database definitions. These data process changes affect the performance measures of a predictive model such as its calibration (how well a model's forecasts reflect the true values)[38] or its discrimination (how well a model can split data into different meaningful categories).[39] The extent of causal predictors used in a model will partly dictate the attrition in calibration and discrimination. A model only associated with the target outcome may suffer degradation in both discrimination and calibration, whereas a more causally founded model may only lose calibration.

During maintenance, both social and technical adjustments are made to optimize and maintain model-related workflows over time. Researchers have not studied this phase as well as the preceding 2 phases. The last portion of the model life cycle is the de-implementation phase.[40–44] In de-implementation, the model is removed or replaced due to a change in workflow, information technology, modeling technique, or some combination of factors.[40] Figure 1 visualizes the stages of the model life cycle—development,[16–21] implementation,[22–28] maintenance,[31–35] and de-implementation[40–44]—along with some example components for each stage. Further details on each component can be found in the accompanying reference in Figure 1.

The diversity of skills needed to develop, deploy, and maintain a model give reason to silo different parts of the model life cycle to different teams. However, compartmentalization may come at the cost of efficiency and efficacy for prognostic models. Prognostic models seek to predict a future patient state, event, or outcomes.[3,45–47] Some prognostic models incorporate the intervention space into the modeling process or forecast nonmodifiable risk (eg, genetic risk).[48–51] Many prognostic models forecast modifiable risk without accounting for interventions.[25,52–55] For example, a model that forecasts the probability of the onset of sepsis in the next 6 hours for a patient on a medical ward is prognostic. After issuing the forecast, clinicians may act by administering broad-spectrum antibiotics or transfer to intensive care. The fundamental purpose of that prognostic model is to improve clinical outcomes, which will change the outcome distribution over time. Deploying a prognostic model into clinical sociotechnical workflows should change behavior. The implicit assumption of the prognostic model is that providing enough forewarning should allow the forecasted outcome(s) to be modified, optimized, or mitigated. This assumption establishes an intimate relationship between the prognostic model and the intervention(s) the model hopes to activate. Despite this interdependence, many models are developed without specifying intervention(s) or without consideration for the implementation phase (eg, readmission risk and suicide attempt risk).[52,56] This divide may help explain the limited impact of readmission risk models on clinical care. Model developers have reported gains in the initial performance of readmission models along with increased provider adoption of readmission risk models,[57,58] yet 30-day readmission rates and age-adjusted mortality rates have stagnated.[59–61] Even fewer models are developed with consideration to the maintenance phase.
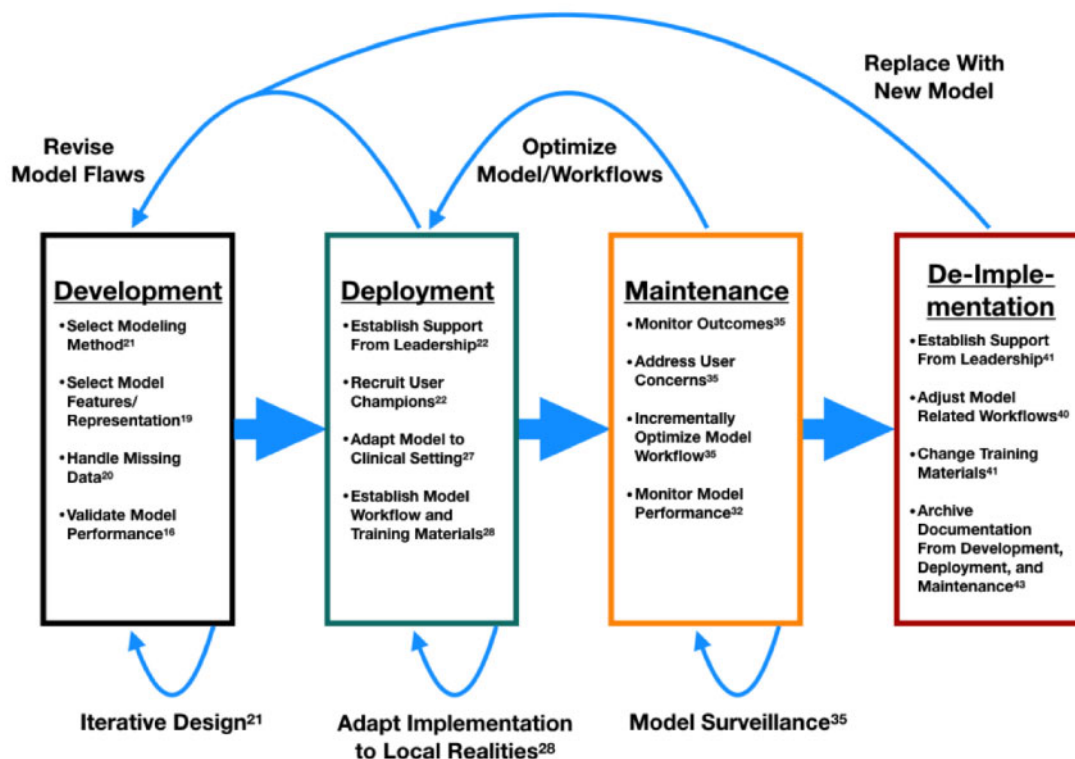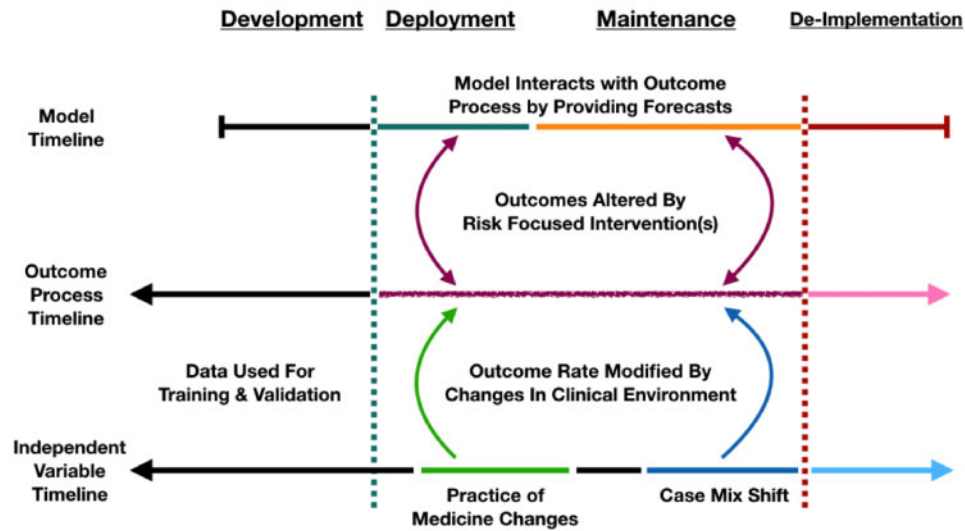

Figure 1. Model life cycle.

Figure 2. Model interactions with outcome.

If end users (eg, patients, clinicians, administrators) act in response to a forecast, then the distribution of the outcome should be altered in some way if the response impacts the outcome. Ideally, prognostic models anticipate and attempt to account for expected changes in the data process due to the fact that they are being introduced. Figure 2 shows the evolution of a prognostic model from observing a clinical process while in development to directly interacting with that process after implementation.

Directly changing the outcome process has strong potential implications for prognostic model maintenance. In the idealized case, end users incorporate the model to great effect. The intervention and model deployment reduce the variance of the outcome, and may shrink the relationship of predictors associated with the outcome toward zero. The magnitude of these effects is directly related to how strongly the intervention impacts the outcome, which can range from very mild impacts (if the intervention is less effective than desired) to very strong ones. The variance of the dependent variable is reduced as the model-directed interventions promote the desired outcomes. With significantly fewer poor outcomes, the distribution of outcomes becomes more concentrated. Actions inspired by the prediction may disrupt the association of predictors to outcomes. This disruption takes place as a result of risk-focused interventions modifying the occurrence of outcomes. For this reason, the successful use of the model results in its own decline in performance. The model is a victim of its own success. Figure 3 demonstrates the results of model refit after the deployment of an intervention on patients at risk in a simple univariate regression simulation. For this example, we assume the presence of a perfect discrimination model. As the intervention gains effectiveness, the slope of the regression line moves toward zero.

This result should be observed in more realistic scenarios in which the prognostic model and the interventions are only moderately effective. As the incidence of the outcome changes in the desired direction, the baseline event rate also shifts, and almost all models are sensitive to changes in overall population outcome event rate. Event rate shifts particularly impact calibration. Under these circumstances, the performance of standard evaluation metrics for calibration, such as log loss, explained variance,[62] or Brier score,[63] and discrimination (sensitivity, specificity, or positive predictive value)[64] will appear to deteriorate. In the idealized case of Figure 3, the model still performs perfectly years later. However, naive assess-

ment of the model without attention to the transformation in the outcome process will result in a suboptimal refit of the model. The refit model will misclassify some patients in need of intervention. Refitting again can lead to the same issue for a different group of patients creating an iterative cycle. Operating the model in a degraded state results in misclassification, potentially leading to safety concerns. This negative cycle may also hasten a model to its de-implementation, diminishing the return on investment of time and capital put into model development, deployment, and maintenance. One could ignore model updating altogether but would still encounter degradation in performance associated with temporal changes of the underlying data process at some point. The same questions of safety and adequate return on investment are also relevant when avoiding model maintenance. There are strategies to mitigate this conundrum by thinking holistically about the model life cycle during model development.

One strategy to avoid this maintenance complication is to include interventions into the model.[1,51] Modeling the intervention space can enable the model to adjust for the actions recommended by model forecasts. This step also allows the model to maintain calibration longer, thereby extending the utility of the model. Including variables on the intervention pathway is not trivial. The interventions may be poorly defined, unrecorded, or not yet implemented. For example, patient-implemented interventions, such as a food or calorie log, may be difficult to find within accessible data sources. Intervention variability, in which only few patients receive the same treatment(s), can lead to biased estimates of effect. For example, a patient at risk for depression may have a large number of preventative options presented to them. Issues of adequate sampling for intervention effect estimation will also plague rare events or outcomes.

If the intervention cannot be directly modeled, informatics practitioners should consider incorporating intervention surrogates. Intervention surrogates are data documenting the workflow associated with an intervention. An intervention surrogate should be correlated with the intervention. While it may not fully capture complex effects of an intervention, it can serve a rough estimate of the intervention's effect. Intervention surrogates can suffer from the same modeling challenges as the interventions themselves.

If modeling interventions directly or through a surrogate is not an option, one can think of the interventions as a latent variable in
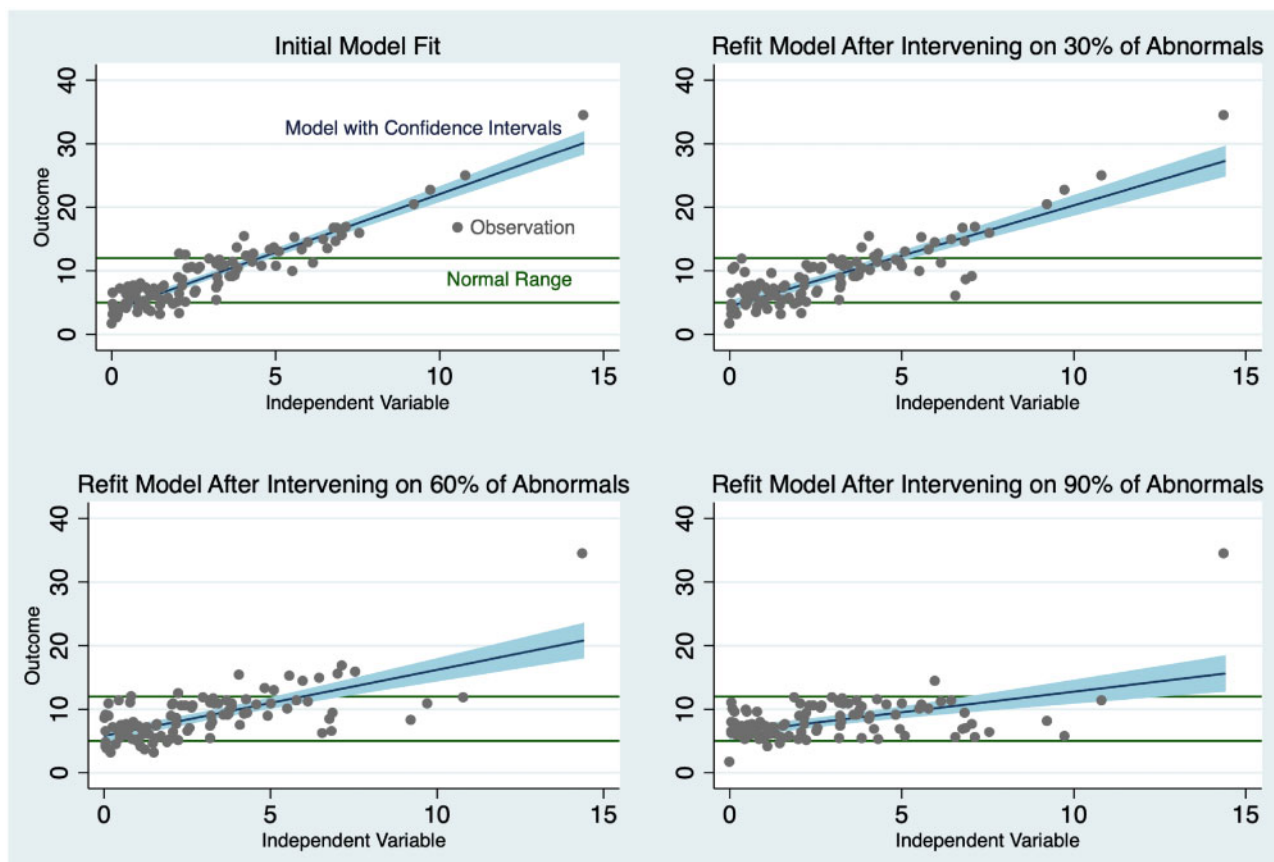
Figure 3. Consequences of model success.

the model. Latent variables are not directly observed but are inferred through a mathematical framework. For example, when building a retrospective cohort to measure the effects on bone density one treatment has against another, there may be confounders that are not present in the available data (eg, whether a subject regularly exercises by jogging or not). These unmeasured confounders will cause variability that cannot be explained using standard modeling techniques. Latent variable models may permit the recovery of other independent variable weights on the outcome but reduce the modeling options available. Generalized mixed effects models, Hidden Markov models and Bayesian frameworks are natural methods for estimating latent variables; however, these methods can add significant complexity to model development or to maintenance efforts.

Model surveillance remains crucial to preserving clinical utility regardless of how one represents interventions in a prognostic model. One cannot begin to diagnosis true and model-mediated performance degradation in a state of ignorance. The efforts to validate and implement a prognostic model into clinical or administrative workflows are not trivial and the potential to cause harm at scale are real.[65–69] It is therefore prudent to design models and their accompanying infrastructure to enable monitoring and maintenance.

Prognostic models produce forecasts on complex and dynamic data processes. Models require surveillance and maintenance over time, so that model developers and implementers can evolve models in concert with the models' nonstationary targets. Failure to monitor prognostic models over time will limit their effectiveness and can in the worst case pose a hazard.[32] However, model surveillance and

updating alone are not enough. If a prognostic model does not incorporate the intervention space directly, through surrogates, or as a latent variable, then model updating may be more frequently required, and may result in exposing more patients to poorly performing model predictions. It is important to keep in mind that the more strongly an intervention impacts the outcome, the more vulnerable prognostic models are to this phenomenon. Prognostic models should be able to mitigate these issues by considering the entire model life cycle as well as the entire outcome process.

FUNDING

AUTHOR CONTRIBUTIONS

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Xiao C, Choi E, Sun J. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* 2018; 25 (10): 1419–28. doi:10.1093/jamia/ocy068.

2. Stead WW. Clinical implications and challenges of artificial intelligence and deep learning. *JAMA* 2018; 320 (11): 1107–8. doi:10.1037/h0030806.

3. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009; 338 (2): b606.

4. Horvitz EF. Data to predictions and decisions: enabling evidence-based healthcare. 2010. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/Evidence_based_healthcare_essay.pdf Accessed February 19, 2019.

5. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*.

6. Hiremath S, Yang G, Mankodiya K. Wearable internet of things: concept, architectural components and promises for person-centered healthcare. In: *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*; 2014: 304–7; Athens, Greece.

7. Wiens J, Shenoy ES. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. *Clin Infect Dis* 2018; 66 (1): 149–53.

8. Ramnarayan P, Tomlinson A, Rao A, Coren M, Winrow A, Britto J. ISABEL: a web-based differential diagnostic aid for paediatrics: results from an initial performance evaluation. *Arch Dis Child* 2003; 88 (5): 408–13.

9. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain. An evolving diagnostic decision-support system. *JAMA* 1987; 258 (1): 67–74.

10. Miller RA, Pople HE Jr, Myers JD. INTERNIST-I, an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med* 1982; 307 (8): 468–76.

11. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc* 2017; 24 (2): 361–70.

12. Parikh RB, Kakad M, Bates DW. Integrating predictive analytics into high-value care: the dawn of precision delivery. *JAMA* 2016; 315 (7): 651–2.

13. Mani S, Ozdas A, Aliferis C, et al. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. *J Am Med Inform Assoc* 2014; 21 (2): 326–36.

14. Pencina MJ, Peterson ED. Moving from clinical trials to precision medicine: the role for predictive modeling. *JAMA* 2016; 315 (16): 1713–4.

15. McKernan LC, Clayton EW, Walsh CG. Protecting life while preserving liberty: ethical recommendations for suicide prevention with artificial intelligence. *Front Psychiatry* 2018; 9: 650.

16. Steyerberg EW, Harrell FE, Borsboom G, Eijkemans MJC, Vergouwe Y, Habbema J. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; 54 (8): 774–81.

17. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016; 74: 167–76.

18. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014; 14 (1). doi:10.1186/1471-2288-14-137.

19. Steyerberg EW, van der Ploeg T, Van Calster B. Risk prediction with machine learning and regression methods. *Biom J* 2014; 56 (4): 601–6.

20. Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis.* New York: Springer; 2015.

21. Armstrong JS. *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Vol. 30. New York: Springer Science & Business Media; 2001.

22. Lorenzi N. Managing change. *J Am Med Inform Assoc* 2000; 7 (2): 116–24.

23. Karsh BT, Weinger MB, Abbott PA, Wears RL. Health information technology: fallacies and sober realities. *J Am Med Inform Assoc* 2010; 17 (6): 617–23.

24. Fleming NS, Becker ER, Culler SD, et al. The impact of electronic health records on workflow and financial measures in primary care practices. *Health Serv Res* 2014; 49 (1pt2): 405–20.

25. Amarasingham R, Patzer RE, Huesch M, Nguyen NQ, Xie B. Implementing electronic health care predictive analytics: considerations and challenges. *Health Aff (Millwood)* 2014; 33 (7): 1148–54.

26. Geissbühler A, Miller RA. A new approach to the implementation of direct care-provider order entry. *Proc AMIA Annu Fall Symp* 1996; 1996: 689–93.

27. Jeffery AD, Novak LL, Kennedy B, Dietrich MS, Mion LC. Participatory design of probability-based decision support tools for in-hospital nurses. *J Am Med Inform Assoc* 2017; 24 (6): 1102–10. doi:10.1093/jamia/ocx060.

28. Elwyn G, Scholl I, Tietbohl C, et al. Many miles to go: a systematic review of the implementation of patient decision support interventions into routine clinical practice. *BMC Med Inform Dec Mak* 2013; 13 (Suppl 2): S14.

29. Bose R. Advanced analytics: opportunities and challenges. *Ind Manage Data Syst* 2009; 109 (2): 155–72.

30. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat Rev Cardiol* 2016; 13 (6): 350–9.

31. Resnic FS, Gross TP, Marinac-Dabic D, et al. Automated surveillance to detect postprocedure safety signals of approved cardiovascular devices. *JAMA* 2010; 304 (18): 2019–27.

32. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. *J Am Med Inform Assoc* 2017; 24 (6): 1052–61.

33. Wright A, Hickman TT, McEvoy D, et al. Methods for detecting malfunctions in clinical decision support systems. *Stud Health Technol Inform* 2017; 245: 1385.

34. Liu S, Wright A, Hauskrecht M. Change-point detection method for clinical decision support system rule monitoring. *Artif Intell Med Conf Artif Intell Med (2005-)* 2017; 10259: 126–35. doi:10.1007/978-3-319-59758-4_14.

35. Bates DW, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc* 2003; 10 (6): 523–30.

36. Minne L, Eslami S, de Keizer N, de Jonge E, de Rooij SE, Abu-Hanna A. Effect of changes over time in the performance of a customized SAPS-II model on the quality of care assessment. *Intensive Care Med* 2012; 38 (1): 40–6.

37. Bosco JLF, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* 2010; 63 (1): 64–74.

38. Gupta HV, Beven KJ, Wagener T. Model calibration and uncertainty estimation. In: *Encyclopedia of Hydrological Sciences*. New York: Wiley; 2005: 2015–31. doi:10.1002/0470848944.hsa138.

39. Linhart H. Techniques for discriminant analysis with discrete variables. *Metrika* 1959; 2 (1): 138–49.

40. Prasad V, Ioannidis JPA. Evidence-based de-implementation for contradicted, unproven, and aspiring healthcare practices. *Implement Sci* 2014; 9: 1.

41. Van Bodegom-Vos L, Davidoff F, De Mheen PJM. Implementation and de-implementation: two sides of the same coin? *BMJ Qual Saf* 2017; 26 (6): 495–501.

42. Norton WE, Kennedy AE, Chambers DA. Studying de-implementation in health: an analysis of funded research grants. *Implement Sci* 2017; 12 (1): 144–13.

43. Chung PWH, Cheung L, Stader J, Jarvis P, Moore J, Macintosh A. Knowledge-based process management—an approach to handling adaptive workflow. *Knowl Based Syst* 2003; 16 (3): 149–60.

44. Parikh M. Knowledge management framework for high-tech research and development. *Eng Manag J* 2001; 13 (3): 27–34.

45. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009; 338 (1): b604.

46. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009; 338 (1): b375

47. Altman DG, Vergouwe Y, Royston P, Moons K. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; 338 (1): b605

48. Collins SP, Lindsell CJ, Clopton P, Hiestand B, Kontos MC, Abraham WT. S3 detection as a diagnostic and prognostic aid in emergency department patients with acute dyspnea. *Ann Emerg Med* 2009; 53 (6): 748–57.

49. Wong HR, Weiss SL, Giulian JS Jr, et al. Testing the prognostic accuracy of the updated pediatric sepsis biomarker risk model. *PLoS One* 2014; 9 (1): e86242. doi:10.1371/journal.pone.0086242.

50. Wong HR, Salisbury S, Xiao Q, et al. The pediatric sepsis biomarker risk model. *Crit Care* 2012; 16 (5): R174.

51. Wong HR, Atkinson SJ, Cvijanovich NZ, et al. Combining prognostic and predictive enrichment strategies to identify children with septic shock responsive to corticosteroids. *Crit Care Med* 2016; 44 (10): e1000–3. doi:10.1097/CCM.0000000000001833.

52. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011; 306 (15): 1688–98.

53. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems. *Ann Intern Med* 2012; 157 (1): 29–43.

54. Kagen D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011; 306 (15): 1688–98.

55. Walsh CG, Sharman K, Hripcsak G. Beyond discrimination: a comparison of calibration methods and clinical usefulness of predictive models of readmission risk. *J Biomed Inform* 2017; 76 (October): 9–18.

56. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull* 2017; 143 (2): 187–232.

57. Brewster AL, Cherlin EJ, Ndumele CD, et al. What works in readmissions reduction how hospitals improve performance. *Med Care* 2016; 54 (6): 600–7.

58. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)* 2014; 33 (7): 1123–31.

59. Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM. Readmissions, observation, and the hospital readmissions reduction program. *N Engl J Med* 2016; 374 (16): 1543–51.

60. Albritton J, Belnap TW, Savitz LA. The effect of the hospital readmissions reduction program on readmission and observation stay rates for heart failure. *Health Aff (Millwood)* 2018; 37 (10): 1632–9.

61. Hall MJ, Levant S, Defrances CJ. Trends in inpatient hospital deaths: National Hospital discharge survey, 2000–2010 key findings. *NCHS Data Brief* 2013; 118: 1–8.

62. DeMaris A. Explained variance in logistic regression study: a Monte Carlo study of proposed measures. *Sociol Methods Res* 2002; 31 (1): 27–74.

63. Murphy AH. A new vector partition of the probability score. *J Appl Meteorol* 1973; 12 (4): 595–600.

64. Pencina MJ, D'Agostino RB. Evaluating discrimination of risk prediction models. *JAMA* 2015; 314 (10): 1063–4.

65. Miller RA. Predictive models for primary caregivers: risky business? *Ann Intern Med* 1997; 127 (7): 565–7.

66. Berger RG, Kichak JP. Computerized physician order entry: helpful or harmful? *J Am Med Inform Assoc* 2004; 11 (2): 100–3.

67. Lepri B, Staiano J, Sangokoya D, Letouzé E, Oliver N. The tyranny of data? The bright and dark sides of data-driven decision-making for social good. In: Kacprzyk J, Ditzinger T, eds. *Studies in Big Data*. Vol 32. New York, NY: Springer; 2016.

68. Dressel J, Farid H. The accuracy, fairness, and limits of predicting recidivism. *Sci Adv* 2018; 4 (1): eaao5580.

69. Miller RA, Schaffner KF, Meisel A. Ethical and legal issues related to the use of computer programs in clinical medicine. *Ann Intern Med* 1985; 102 (4): 529–36.