

Issue BRIEF

✓ CHECKLIST

Is a predictive model an appropriate tool?

- Do you have a measurable outcome that you are trying to achieve?
- Do you have an intervention that can help achieve this outcome?
- Is a targeted approach more appropriate than a universal approach for this intervention?
- Does the predictive modeling effort meet the following technical conditions?
 - Adequate predictive performance that is better than alternative targeting options
 - Adequate statistical power
 - Availability of high-quality data and analytic capacity



GEORGETOWN UNIVERSITY
 McCourt School of Public Policy

FEBRUARY 2017

MATHEMATICA-MPR.COM

Lindsey Leininger, Mathematica Policy Research; and Thomas DeLeire, Georgetown University

Predictive Modeling for Population Health Management: A Practical Guide

Predictive modeling is frequently used in population health management programs to stratify populations by their risk of a poor health care outcome. This brief outlines scenarios for which a predictive modeling application is likely to be appropriate and describes key practical considerations for implementation.

WHEN IS PREDICTIVE MODELING APPROPRIATE?

Four steps will help you assess whether the development and deployment of a predictive model is an appropriate tool for a particular population health context: (1) define the specific health care outcome that you're trying to achieve; (2) identify an existing or a new intervention to help achieve the desired health care outcome; (3) identify the appropriateness of a targeted versus universal intervention approach; and (4) if using a targeted approach, decide whether to implement a predictive model for identifying the groups to be targeted.

Define the specific health care outcome

First, it's important to identify the ultimate outcome you want to achieve. In population health management, often the primary goal is to reduce hospital-based care such as emergency department (ED) visits, inpatient stays, and readmissions. Other example objectives include reducing patient absenteeism and medication nonadherence. To be appropriate for a targeted approach supported by a predictive model, the key outcome must be a measured one (or measurable in the future).

Identify an intervention: existing or new

Various types of case management and care coordination typically serve as bedrock interventions for population health management. It's essential that the intervention is well designed and appropriately tailored to the population, that it exhibits a compelling theory of change, and that it is accepted by clinicians and other potential end users. Without an effective intervention, the best, most sophisticated risk stratification system—whether it is built on a predictive model or another mechanism—cannot improve population health outcomes. *Developing and deploying a predictive model would not be appropriate in the absence of such an intervention. Thus, intervention design and implementation must be the highest-priority focus.*

Identify the appropriateness of a targeted versus universal intervention approach

- Is a targeted approach appropriate from a **clinical perspective**? Many population health management interventions are effective for high-need populations only. For example, a review of Medicare care coordination programs reveals that their success is limited to populations with relatively high

health care needs.¹ Typically a targeted approach to implementation is better than a universal approach in situations where likely intervention effectiveness varies substantially across the patient population.

- Is it appropriate from a **risk distribution perspective**? A targeted approach is appropriate for populations that exhibit a wide distribution of risk, such as a general pediatric panel in which some identifiable groups of children face a high risk of potentially preventable ED utilization (whereas most face a low risk). In contrast, a universal approach is preferred for populations with uniformly high levels of underlying risk, such as homeless individuals who have multiple chronic conditions. If all (or nearly all) patients are at high risk of a poor health outcome, a targeted approach may not be appropriate. The appropriateness—and indeed importance—of targeting increases in tandem with the spread of underlying risk in the patient population.
- Is it appropriate from a **resource perspective**? Often the adoption of a targeted approach is ideal or necessary due to resource constraints. Business case modeling provides guidance with respect to the likely intervention return on investment (ROI) across different patient risk tiers.² The resulting business case estimates help illustrate whether targeting, and which type of targeting, is most likely to achieve a positive ROI.
- Is it appropriate from an **end-user perspective**? It is critical to assess clinician willingness to implement a targeted approach. In some contexts, clinicians may advise against such an approach, instead preferring to adopt a universal intervention. For example, numerous clinicians participating in a Medicaid-supported targeted case management program for high-risk pregnant women in Wisconsin felt an ethical obligation to provide intervention services to all patient populations.³ Simply stated, they believed that their patient panels were universally high need. Acquiring the input and support of clinicians in implementing a targeted approach is key to ensuring intervention success.

Decide whether to implement a predictive model

For a targeted approach, the following considerations are key in deciding whether to use a predictive model to identify groups to be targeted:

- **Likely predictive performance.**
A predictive model is ineffective in targeting appropriate interventions if it cannot accurately classify patients by level of risk of a poor health outcome. What is the likely classification accuracy of a predictive model? Look to both scientific literature and similar real-world examples to assess the potential performance of a predictive model; the existence of models that perform well in similar contexts provides reassurance regarding the success of a similar application in yours. Classification accuracy is typically assessed using the *c-statistic*. By far the most universally cited statistic for predictive models, the *c-statistic* is the easiest piece of information to find in a literature search or environmental scan. The *c-statistic* ranges from 0.5 to 1, with 0.5 indicating that the model is no more accurate than a random coin flip in classifying high- versus low-risk individuals, and 1 indicating that the model correctly classifies high- versus low-risk individuals in every case. Although there is no uniformly accepted *c-statistic* threshold for predictive performance, a commonly accepted rule of thumb is 0.7. Some common population health outcomes meet this performance threshold—for example, predictive models of high ED utilization generally range from 0.7 to 0.8—while others, like models predicting 30-day readmissions, typically do not.⁴
- **Statistical power.**
Does your patient population provide a sufficient number of observations to develop and validate a predictive model? As with the *c-statistic*, there are no formal criteria for assessing likely statistical power; however, there are rules of thumb.⁵ It is important to be aware that when a patient population is too small to support a model containing the appropriate number of predictors, it may be tempting to target interventions using the results and weights from an existing predictive model developed and validated on a different population. *Typically, this approach is inadvisable.*⁶ Even a statistical model with the same outcome and predictor variables can yield very different weights and aggregate risk scores

in data sets from different patient populations. Although it is appropriate—and indeed advisable—to estimate existing statistical models using your patient population data, it is rarely appropriate to calculate aggregate risk scores using weights estimated from existing models on other patient populations. Accordingly, if your population is too small to support its own model estimation, you’ll likely want to abandon predictive modeling in favor of a different targeting approach or a universal approach.

- **Data availability and analytic capacity.** Implementing a predictive model for risk stratification purposes requires *access to needed data* and *access to analytic capacity* to manipulate the data and estimate the model. Are the requisite data for model estimation available and easily accessible? Do you have staff members who know how to extract, clean, and analyze data? If not, do you have access to a contractor or another trusted source that can perform these functions for you? Are the costs associated with developing or contracting for the capacity to implement a predictive model worth it from a resource perspective?
- **Alternatives to predictive modeling.** Which targeting alternatives are in practice currently or potentially available to you? Common examples include physician referrals, self-identification, and case manager judgment. It’s important to assess whether a predictive model is likely to stratify risk more effectively than the alternatives. Compelling research demonstrates that even the simplest predictive model can often outperform expert judgment.⁷ This “relative to what?” question is an important benchmark to keep in mind, and is often the reason lower-performing predictive models in some contexts—such as those for readmissions—may be the most appropriate stratification tool available.

IMPORTANT CONSIDERATIONS FOR PREDICTIVE MODELLING IMPLEMENTATION

After determining that predictive modeling is worthwhile, it is important to allow sufficient time for procuring and cleaning the requisite data as well as developing, validating, and estimating the statistical model. Notably, if the underlying databases supporting the model are not currently

designed for similar analytic purposes (which is often true for electronic health records and claims data), the data cleaning process will take the most time (by far) of these tasks. In contrast, once the data have been cleaned and organized, subsequent statistical work can be done relatively quickly in many commonly used statistical packages (for example, Stata and SAS), resulting in the final output of the modeling effort: predicted risk scores for each patient in the panel.⁸

Two big decisions must be made after receiving the patient-level predicted risk scores: (1) where to draw the line in the predicted risk distribution for differentiated intervention, and (2) whether to intervene/surveil (differentially) across all risk tiers or in the highest tier only.

Where to draw the line in the predicted risk distribution

This is a *policy* question that requires subject matter expertise, not a statistical question. The most important organizing principle to keep in mind is that predictive modeling is a form of health care technology, and like any other technology it is *not* a (stratification) silver bullet. It’s best to treat the model estimates as a clinician would treat results from a *screening* test, not a definitive diagnostic test. Risk scores resulting from population health applications are associated with an appreciable amount of uncertainty, reflecting a baseline irreducible uncertainty associated with predicting future health care needs.

The guiding criterion for choosing the threshold for differentiated intervention should be your *relative preferences regarding false positives versus false negatives*. In this context, false positives would suggest implementing a high-risk intervention for a patient who is not, in fact, high risk. (That is, the patient would be erroneously classified as high risk.) False negatives would suggest not providing the appropriate high-risk intervention to a high-risk patient, because that patient would be erroneously classified as low risk. There is *always* a tradeoff between false negatives and false positives in risk stratification; with the given data and model, *there is no way to decrease both simultaneously*.⁹ If you decrease the false positive rate, the false negative rate will increase. To minimize the false negative rate, you would need to categorize as high risk many individuals with relatively low predicted risk

scores. The implication is that more individuals would be flagged as high risk, requiring greater program resources. Moreover, many individuals would be incorrectly flagged as high risk, resulting in an intervention that is administered to many individuals who don't need it.

Alternatively, an attempt to decrease the false positive rate would exclude individuals with all but the highest predicted risk scores. The implication is that fewer individuals would be flagged as high risk, requiring fewer program resources. However, many individuals would be *incorrectly* flagged as low risk; thus the intervention would fail to reach many needy patients who would benefit from it.

Whether to intervene/surveil (differentially) across all risk tiers or in the highest tier only

Targeted population health management initiatives typically apply one of two approaches:

- The highest-risk group receives the intervention, and none of the other tiers receive any intervention or further surveillance (high-risk-only approach).
- Each risk group receives a differentiated amount of services (and/or further surveillance), increasing in intensity for higher-risk tiers (differentiated approach).

Here again, the tradeoff between the false positive and false negative rates is an important consideration. An intervention designed to minimize the false positive rate will, by construction, have a high false negative rate, *excluding many needy patients*. Accordingly, it is advisable to either implement a differentiated approach or pair a high-risk-only approach with a system of continued active or passive surveillance of the predicted low-risk population. For example, the highest-risk group might receive resource-intensive nurse home visits while the lower-risk groups receive occasional phone calls from a case manager. Post-risk stratification, allowing for clinician referral in the intervention is another surveillance technique that can identify needy patients missed by the predictive model. In sum, it is key to implement a way to identify the needy patients from the large number who will invariably be misclassified as low risk.

In contrast, a strategy that minimizes false negatives *will include many patients who do not need the intervention*. In keeping with clinical medicine, which almost always seeks to minimize false negatives in screening applications, the best practice for patients receiving a high-risk score typically involves further follow-up—often performed by a physician or case manager—to provide a more accurate diagnostic assessment of need. In other words, it is important to filter out the large number of patients that the model will misclassify as high risk. This second line of diagnostic filtering is particularly important for high-risk-only approaches that employ expensive interventions and/or interventions that have demonstrated effectiveness in high-need populations only.

OTHER HELPFUL POINTS TO CONSIDER

Be mindful about the *time* aspect of risk.

Individuals' underlying risk of a poor health outcome can change considerably over time; thus stratification approaches need to reflect this dynamism. Specifically, risk scores should be reevaluated at an interval that makes sense from a clinical or policy perspective. A calendar year, which dovetails with the duration of many health insurance contracts, is a frequently used interval; however, other intervals may be appropriate depending on the intervention and the population served.

Many individuals whose health care needs are elevated during one period have lower health care needs in another. Even without a population health management intervention, many high-risk individuals will exhibit lower health care utilization at times in the future. This statistical phenomenon is called “regression to the mean.” Note that regression to the mean coexists with the fact that prior health care utilization is almost always the best predictor of future utilization in predictive models. These two phenomena are not mutually exclusive. It's important to understand that some people in the high-risk population will likely require a shorter intervention while others will require a longer one; however, it's typically very difficult to predict into which group a given high-risk individual will fall. This uncertainty underscores

the need to reassess risk scores on an ongoing basis, as discussed previously. Another important implication of regression to the mean is that comparisons of high-risk individuals' outcomes before and after an intervention will yield incorrect inferences about the intervention's success, thus overstating the impact. Accordingly, it is crucial to find a comparison group to serve as a benchmark for the high-risk intervention group's pre- and post-outcomes. This need for a credible external benchmark highlights a general rule: that having a valid and useful predictive model does not mean that subsequent impact evaluation will be straightforward. Predictive models and evaluation methods are distinct tools; however, to be successful, both must be designed and implemented prior to program launch. Also, both can be deployed in a continuous quality improvement cycle, using rapid cycle evaluation to assess the intervention's effectiveness and continually refine the predictive model to improve targeting.

What should you do if you lack historical data on your patient populations at program inception? It's ideal to build a predictive model using historical data from your patient population. However, organizations often lack such data at the time of intervention rollout, intending to rely on data gathered prospectively to populate future predictive models. There are two options for initial stratification in this scenario: until the organization's own data can support estimation and validation of weights for the patient population to interest, (1) use coefficients/weights from a similar predictive model implemented on a different patient population, or (2) use clinical/policy judgment for the initial period.

As discussed previously, the first option is not recommended because it would likely misfire; weights derived from similar predictive models for different patient populations can be appreciably different. Similarly, the use of clinical/policy judgment may also misfire—predictive models can often outperform expert judgment. The second option is often better because it tends to have more end-user acceptance. In fact, if the first approach is implemented and fails, it can engender skepticism about predictive modeling in general.

CONCLUSION

This brief discusses key considerations in the decision to adopt a predictive modeling approach for population health management interventions. It also provides guidance for stakeholders who decide to implement a high-risk approach supported by a predictive model. Throughout, we stress the importance of pairing statistical modeling with programmatic expertise; indeed, one of our primary objectives in this brief is to demonstrate that the ultimate success of predictive modeling in population health management requires much more than statistical sophistication. A successful strategy depends upon identifying (1) a precisely defined health care problem, (2) a promising intervention, (3) the appropriateness of a targeted approach in implementing the intervention, and (4) sufficient data and analytic capacity to develop and validate a predictive model. Note that only one of these four steps is statistical in nature.

Once the decision has been made to build a predictive model, it is critical to remember that the resulting stratification is akin to a screening test, not a diagnostic one, and often requires further risk assessment and/or monitoring. Moreover, there is always a tradeoff between false positives and false negatives in the decision about where in the predicted risk distribution to intervene. Finally, since underlying risk of elevated health care need is dynamic, it's important to reassess risk at clinically relevant intervals and avoid evaluating interventions based on simple pre- and post-comparisons of the high-risk intervention group.

ENDNOTES

¹ Peikes, D., A. Chen, J. Schore, and R. Brown. "Effects of Care Coordination on Hospitalization, Quality of Care, and Health Care Expenditures Among Medicare Beneficiaries: 15 Randomized Trials." *JAMA*, vol. 301, no. 6, February 11, 2009, pp. 603–18.

² For a nice example see Billings, J., and T. Mijanovich. "Improving the Management of Care for High-Cost Medicaid Patients." *Health Affairs*, vol. 26, no. 6, November–December 2007, pp. 1643–54.

³ Friedsam, D., L. Leininger, and K. Voskuil. "Evaluation of Medicaid Medical Homes for Pregnant Women in Southeast Wisconsin. Final Report." Madison, WI: University of Wisconsin Population Health Institute, March 2016.

⁴ ED visits: Wherry, L. R., M. E. Burns, and L. J. Leininger. “Using Self-Reported Health Measures to Predict High-Need Cases Among Medicaid-Eligible Adults.” *Health Services Research*, vol. 49, Suppl 2, December 2014, pp. 2147–72.

Readmissions: Kansagara, D., H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, and S. Kripalani. “Risk Prediction Models for Hospital Readmission: A Systematic Review.” Portland, OR: Evidence-Based Synthesis Program (ESP) Center, Portland VA Medical Center, 2011.

⁵ The associated technical details are beyond the scope of this brief; a nice related resource is Harrell, F. E., K. L. Lee, and D. B. Mark. “Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors.” *Statistics in Medicine*, vol. 15, no. 4, February 28, 1996, pp. 361–387.

⁶ For exceptions see Shwartz, M., and A. S. Ash, “Empirically Evaluating *Risk Adjustment Models*.” In *Risk Adjustment for Measuring Health Care Outcomes*, edited by Lisa Iezzoni. Chicago, IL: Health Administration Press, 2013.

⁷ Kahneman, D. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux, 2013.

⁸ For an accessible guide to statistical modeling for predictive purposes, see Steyerberg, E. W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer, 2009.

⁹ Adopting a model with better predictive performance (with a higher *c*-statistic) would allow you to reduce both false positives and false negatives. In practice, this typically requires access to new, richer data.

For more information, contact Lindsey Leininger at lleininger@mathematica-mpr.com or Thomas DeLeire at thomas.deleire@georgetown.edu

