



REFLECTIONS
ON
METHODOLOGY

SEPTEMBER
2017

MDRC's Approach to Using Predictive Analytics to Improve and Target Social Services Based on Risk

By *Kristin E. Porter, Rekha Balu, and Richard Hendra*

This post is one in a series highlighting MDRC's methodological work. Contributors discuss the refinement and practical use of research methods being employed across our organization.

Across policy domains, practitioners and researchers are benefiting from a trend of greater access to both more detailed and frequent data and the increased computing power needed to work with large, longitudinal data sets. There is growing interest in using such data as a case management tool, to better understand patterns of behavior, better manage caseload dynamics, and better target individuals for interventions. In particular, predictive analytics — which has long been used in business and marketing research — is gaining currency as a way for social service providers to identify individuals who are at risk of adverse outcomes. We model the experiences of individuals whose outcomes are known in order to predict outcomes for others.

MDRC has multiple predictive analytics efforts under way, which we summarize below while highlighting our methodological approach. Some of these efforts focus on applying predictive analytics to produce new information that is meaningful and actionable for practitioners. Other efforts focus on assessing whether predictive analytics is an appropriate and feasible approach for a given organization and evaluating the implementation, value, and limitations of existing predictive analytics tools.

At MDRC, we are using predictive analytics to **estimate an individual's likelihood of achieving or not achieving key outcomes** (such as completing a program milestone, finding employment, or reading at a proficient level).

- We use *machine learning algorithms*, which allow us to extract information from a large number of measures (perhaps hundreds) and thus capture wide-ranging risk factors as well as possible heterogeneity in risk factors across clusters of individuals. Many of the algorithms also let the data determine what form a statistical model should take (rather than the researchers specifying a functional form).
- Moreover, we use *ensemble learning*, which allows us to compare multiple machine learning algorithms and other modeling approaches and select the best one based on its predictive performance in *new* samples. With ensemble learning, we can also *combine* information from different model-building approaches — thereby employing an algorithm of algorithms. Evaluating models with data that were not used to train the models helps avoid overfitting, in which the models do not adequately capture underlying relationships in the data and therefore do not generalize well to new data. By using ensemble learning, we are able to prespecify both models that incorporate substantive knowledge and hypotheses about the importance of predictive factors *and* models that are built by algorithms. We can then assess the relative performance of the different types of modeling approaches. More details about our predictive modeling approach can be found in our [Primer for Researchers Working with Education Data](#).

NEW YORK

16 East 34th Street
New York, NY 10016
Tel: 212 532 3200

OAKLAND

475 14th Street
Suite 750
Oakland, CA 94612
Tel: 510 663 6372

WASHINGTON, DC

1990 M Street, NW
Suite 340
Washington, DC 20036

LOS ANGELES

11965 Venice Boulevard
Suite 402
Los Angeles, CA 90066

www.mdrc.org



In recent and ongoing researcher-practitioner partnerships with service providers, we are incorporating our predictive analytics approach into institutions' caseload management systems and **continuous improvement processes**. For example, with school districts and education support organizations, we are using rich, student-level longitudinal data to predict low-income students' school performance outcomes and achievement of milestones toward high school graduation. And, in a partnership with the [Center for Employment Opportunities](#), which helps people who have been incarcerated find and keep jobs, we are using longitudinal administrative data to predict participants' risk of not reaching program milestones and completion. The predictions can be used to target and refine program services, and these improvements can be evaluated with rapid-cycle randomized controlled trials.

- Our methodologists strongly believe that working in *close partnership with service providers* is critical not just for maximizing the usefulness of the analysis but also for maximizing the *accuracy* of predictions. A deep understanding of the data and the contextual factors behind the data is often more important to a robust predictive model than the cutting-edge methods described above. Practitioners' knowledge helps us select and create measures that are most predictive of the outcomes of focus and that have consistent meaning over time. The resulting predictions are expected to lead to better identification of at-risk participants and better tailoring of services.

With other institutions, we are actively exploring whether predictive modeling is the right type of analysis for their goals and questions. If so, we then investigate their **readiness**, the extent to which the existing organizational capacity and data systems allow for rapid and iterative predictive modeling of milestones of interest. Our [practitioner brief](#), which describes the value of predictive analytics for institutions that support young people in and out of school, outlines our approach to determining readiness. The brief also contrasts our method with the prevailing approach for estimating risk in programs that work with young people, which relies on "indicators" rather than predicted risk levels.

Finally, MDRC is also studying the **implementation and use of existing, validated predictive analytics tools**. For example, our [process and impact evaluation of New York City's pretrial Supervised Release program](#) includes an investigation of the program's risk assessment tool, which uses predictive analytics to estimate a defendant's risk of felony rearrest during the pretrial period. Lessons learned from the actual adoption of tools such as this one can provide valuable information for methodological consideration. For example, to what extent do practitioners value or even require transparency in the model behind a risk assessment tool (in terms of the measures that make up the tool and the method by which the measures are combined), versus the potential gains in accuracy offered by complex algorithms? Also, to what extent does predictive analytics improve human decisions? Does predictive analytics reduce discrimination relative to human decisions — or does this approach instead exacerbate discrimination already ingrained in measures in the data?

MDRC's framework for predictive analytics includes a set of decision rules for processing and cleaning data in ways that are optimal for the prediction objective. Moreover, the framework includes guidelines for interpreting prediction results — including cautions about the limitations — and for translating the results into decisions that can drive an institution's continuous learning process. The framework is designed to foster an ongoing process that updates models and results as new data are collected, so that individuals' risk estimates can also be updated over time.

In sum, predictive analytics allows us to combine the considerable power of machine learning methods and the critical expertise of analysts and practitioners to help social service providers make the best use of their resources.