Actionable Evidence Initiative Case Study

Unlocking Real-Time Evidence for Practitioners

How Evaluation and Data Analytics Are Generating On-Demand, Actionable Evidence for Front-Line Practitioners at First Place for Youth and Gemma Services

Peter York, BCT Partners

July 2021

PROJECT EVIDENT

Support provided by:

BILL & MELINDA GATES foundation

The Actionable Evidence Initiative

Led by Project Evident with funding from the Bill & Melinda Gates Foundation, the Actionable Evidence Initiative seeks to understand and remove barriers to building evidence that is equitable, useful, credible, and relevant for practitioners as they aim to improve the outcomes of students who are Black, Latino/a/x, or experiencing poverty. Please visit https://www.projectevident.org/actionable-evidence to learn more, join our network, and find partners interested in working together on actionable evidence solutions.

Actionable Evidence in Education Cases

This case is one in a series commissioned by the Actionable Evidence Initiative in 2020 and 2021. (Cases are published on the Project Evident website.) The series illustrates how researchers, evaluators, practitioners, funders, and policymakers across the country are exemplifying principles of the Actionable Evidence framework. It profiles a range of settings, actors, learning questions, methods, and products, unified by a commitment to practitioner-centered, timely, practical, equitable, and inclusive evidence building. Each case describes the origins, development, and results of a research or evaluation project, along with the authors' reflections on their experiences. Our hope is that these cases will provide both inspiration and practical guidance for those interested in generating and using evidence that leads to better and more equitable outcomes for youth and communities.

Contents

Actionable Evidence in Education: Unlocking Real-Time Evidence for Practitioners

Peter York

Executive Summary

First Place for Youth helps youth who have aged out of the child welfare system build the skills they need to make a successful transition to self-sufficiency and responsible adulthood, and Gemma Services is a social service agency that operates a long-term residential psychiatric care program for youth. A few years ago, both organizations realized that their administrative data systems, while extensive, generated little data that could be used by front-line practitioners as they worked directly with youth and families.

To address this limitation, First Place and Gemma partnered with BCT Partners, an evaluation and data science firm, to institute "precision analytics," which trains machine learning algorithms to build predictive, prescriptive, and evaluation models that can be technologically implemented to generate actionable evidence, on demand. Using program administration and case assessment data, this process learns from differences in outcomes among similar youth who experienced different treatment patterns. The goal of this process is to understand what has worked for specific participant populations in the past and then to translate those findings into actionable information for front-line staff so that they can connect each youth with the services and supports that are most likely to lead to success.

This case describes the steps the partners at both organizations took to develop their learning systems and user dashboards; the resources and capacity required, including the roles of evaluators, practice experts, data scientists, and practitioners; and the resulting changes in practice for front-line staff, managers, and leadership. It also addresses steps the partners took to minimize bias in the development of algorithms.

About the Project

Origins

The leaders of two youth-serving organizations, First Place for Youth and Gemma Services, believed strongly in the need to rigorously evaluate their programs so they could produce the kind of evidence that would advance their programs and practices, as well as hold themselves accountable to achieving positive client outcomes. Both organizations had considered conducting evaluations using traditional randomized controlled trials and quasi-experimental evaluation designs to ensure that their programs in fact made a difference. Instead they chose to invest in program administration data systems that would serve to assess, monitor, and evaluate the outcomes of every client throughout their program experience. These data collection efforts included gathering pre-test/post-test client data using scientifically validated assessment instruments, and Gemma Services even engaged a staff member to conduct follow-up interviews and surveys with clients three months and six months after they had been discharged to evaluate their long-term outcomes.

Both organizations reached the conclusion that, while these data systems served an important program administration purpose, including being able to report to their funders on service outputs and costs, they weren't meeting their evidence generation needs. This was especially the case when it came to practitioners and clients. Practitioners (front-line staff) were the principal data collectors, spending hours every week gathering information and assessments from clients and inputting data into the system. However, they received no evidence in return on which they could act to learn about and strengthen their program planning and engagement.

Figure 1: About Gemma Services and First Place for Youth

Gemma Services	**Annual Budget: ~$8-10 MM**
Overview: Gemma Services is a multi-program social service agency providing behavioral health, education, and prevention services for children and families experiencing emotional and behavioral difficulties. Gemma Services Psychiatric Residential Treatment Facility (PRTF) serves a maximum of 72 children at a time and generally an average of 68. The PRTF is one step below a locked psychiatric hospital in the continuum of restrictiveness. The overall objective of the program is to decrease children's distress and help improve their lives, ideally enabling them to return to live successfully in a home or community setting.	
Youth Served: ~100-110 in the PRTF annually. Youth range from six to 14 years old and primarily come from the 5-county Southeastern Pennsylvania region as well as surrounding counties. The youth must meet medical necessity to be referred by a psychiatrist or medical professional for an intensive residential placement due to their present difficulty keeping themselves or others safe while living in a community setting. They are experiencing significant emotional and behavioral difficulties and may have experienced traumas such as physical, sexual, or emotional abuse; neglect; exposure to domestic or other violence; incarceration of a parent; and more. Approximately one-third of the children served by the program have no family involved in their lives. Often, they are referred to a PRTF after long and unsuccessful involvement within the mental health system.	

First Place for Youth	**Annual Budget: ~$25 MM** (FY2019)
Overview: Founded in 1998, First Place for Youth™ seeks to help foster youth build the skills they need to make a successful transition to self-sufficiency and responsible adulthood. Through the flagship program, My First Place™ (MFP), and other supporting programs, First Place supports at-risk foster youth at a critical time in their lives when they need to learn to support themselves. Since its founding, First Place's work has been driven by two goals: provide results-driven direct service to young people who need its help; and change public policies to improve the lives of as many transition-age foster youth as possible.	
Youth Served: 1,583 across all programs in 2020**.** First Place currently serves youth in six California counties (Alameda, San Francisco, Contra Costa, Santa Clara, Solano and Los Angeles), and has affiliate partners in New York, Massachusetts, Mississippi, and Ohio. In FY2020, First Place and its affiliates served a total of 1,583 youth and housed 597 youth in MFP. Through MFP, First Place serves young people (18-24) who are or were in the child welfare or probation systems. The typical foster youth entering the program has been exposed to multiple traumatic events and was removed from their home at 12 years old. They were then placed into and removed from five different foster homes or placements. Most of First Place's young people have grown up in the system, having been in the custody of the state for nearly eight years on average. Among youth entering the program in FY20, 49% were unemployed, 21% did not have a high school diploma/GED, 73% were not stably housed, 41% had been arrested, and 20% had children.	

After years of implementing their data systems, leaders from First Place for Youth and Gemma Services became aware of analytic studies in the fields of justice[1] and child welfare[2] that used program administration data to show how predictive real-time actionable evidence could be made available to their front-line practitioners, on demand. These studies found that machine learning algorithms could be used to rigorously evaluate social programs, at least quasi-experimentally. These analytic studies combined social science and data science methods and tools to evaluate programs, as well as to make case-specific predictions and recommendations that reminded the leaders of how Netflix and Amazon can make tailored and precise recommendations that predict a positive outcome (i.e., enjoyment) if a person watches a specific movie or reads a specific book. These juvenile justice and child welfare data studies found unique groups of youth and identified the tailored pattern of program design features that maximized the odds of achieving the desired outcome.

> "The for-profit sector has long embraced cutting-edge technology and tools to capture and use data to better serve their customers… However, these tools, along with the talent and services to guide those who use them, have largely not made it into the hands of the nonprofit sector, where the outcomes hold enormous consequences not just to individuals but to the well-being of our communities.
>
> Ben Berres
> Solving the "Last Mile" Problem of Technology, Tools, and Talent in the Social Sector

[1] Using Predictive Analytics and Machine Learning to Improve the Accuracy and Performance of Juvenile Justice Risk Assessment Instruments
[2] Predictive and Prescriptive Analytics, Machine Learning and Child Welfare Risk Assessment: The Broward County Experience

First Place for Youth and Gemma Services' leaders learned how machine learning algorithms were being trained to not only determine the odds of success for different segments of the population, but to find the exceptions to the rule – those cases beating the odds, or the 'positive deviants' – to discover what they experienced that made the difference.

Around 2018, the leaders of both First Place for Youth and Gemma Services decided that they wanted to try adopting these analytic methods and the tools of data science, including predictive and prescriptive analytics, so their practitioners could access information that could help them increase the odds of success for each and every youth they served. When these organizations heard about the juvenile justice and child welfare examples above, they realized that it was possible to use the program administration data they were already collecting to finally provide their practitioners with actionable evidence in return for all of their efforts. They sought to build predictive and prescriptive models with the primary users being front-line practitioners. Their goal was to develop a technological solution that would put on-demand actionable evidence in the hands of their practitioners on a daily basis.

Precision Analytics to Generate Actionable Evidence

The precision analytics approach that First Place for Youth and Gemma Services adopted, with support from BCT Partners, is designed to meet different needs than traditional summative evaluation designs such as experimental randomized controlled trials (RCTs) and quasi-experimental matched comparison group studies. While such designs are useful in examining the overall effectiveness of a program, practitioners who are making choices about how to treat or serve the client they are meeting right now need more contextualized and precise information.

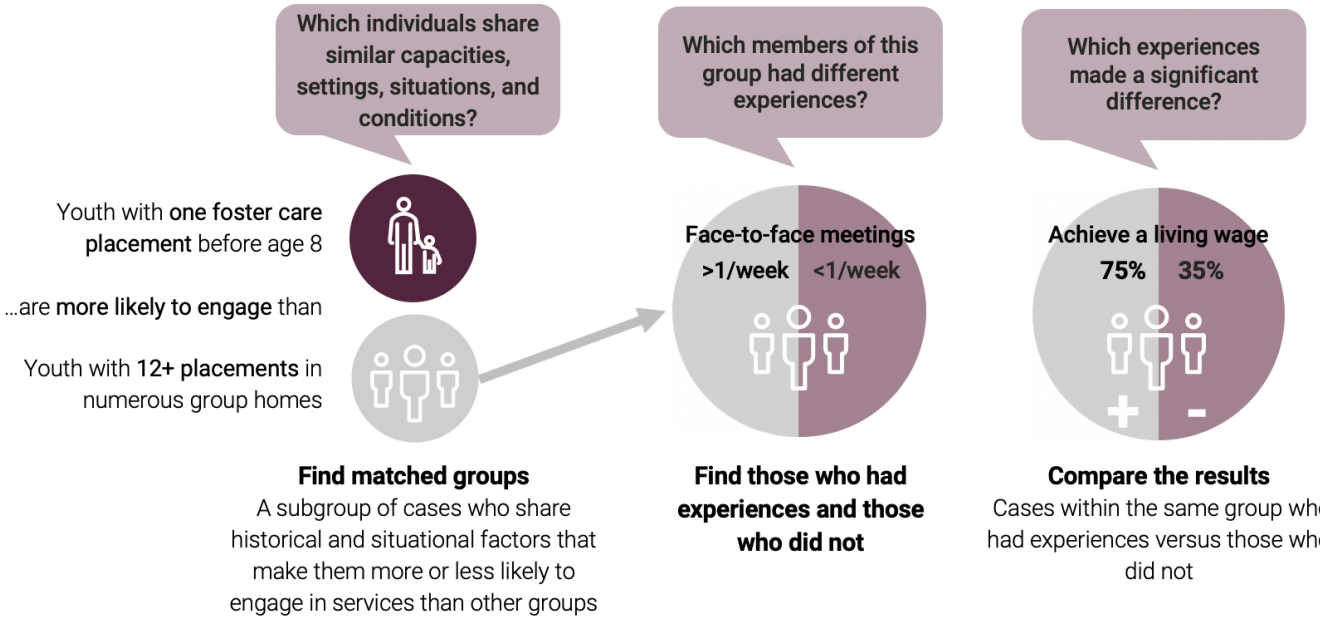
> "To get better results in this complex world, we must be willing to shake the intuition that certainty should be our highest priority.
>
> Lisbeth Schorr
> Reconsidering Evidence: What It Means and How We Use It

The data science and machine learning algorithms used in the private sector to predict odds of success and make tailored recommendations hold promise for addressing this need in the social sector. Evaluators and data scientists at BCT Partners have developed, tested and published reports[3] on an innovative approach – Precision Analytics (PA) – which trains machine learning algorithms to build predictive, prescriptive, and evaluation models that can be technologically implemented to generate actionable evidence, on demand. The PA process uses program administration and case assessment data to find matched comparison groups of like cases, based on their historical and baseline strengths, challenges and environmental context. Then, machine learning algorithms are trained to find all the different ways in which children and families in the same

[3] Scattergood Foundation funded the publication of the Gemma Services (formerly known as the Silver Springs-Martin Luther School) case study, A New Way to Use Data: Precision Care for Better Outcomes in Psychiatric Residential Treatment for Children; the Rockefeller Foundation published the paper, Measuring Results and Impact in the Age of Big Data: The Nexus of Evaluation, Analytics, and Digital Technology, which highlighted examples of the precision modeling approach.

group were treated differently; these are the naturally occurring real-world "experiments" in which the same types of children were unconsciously being assigned to different treatments. Algorithms find these counterfactual treatments within a very similar group of cases, determine what pattern of treatments achieved the greatest comparative gains, and report on these "ideal treatments," including tests of significance and the effect sizes of the program model. Expert practitioners then develop the language and narratives that translate these patterns into useful information for front-line staff, including on-demand predictive and prescriptive insights presented in dashboards, reports, and visualizations, while also providing leaders and funders with up-to-date, precise, rigorous evaluation findings about how many lives are being attributably improved.[4]

Figure 2: Precision Modeling: Finding natural experiments in history



It's the machine learning version of propensity score matching.

[4] This approach is aligned with the application of the causal analytic techniques put forth by machine learning pioneer, Turing Prize winner, and causal modeling author and expert Judea Pearl in The Book of Why: The New Science of Cause and Effect.

Beginning the Data-Driven Actionable Evidence Journey

Engaging Practitioners

The leaders of First Place for Youth and Gemma Services knew that successful adoption of precision analytics would require the buy-in and support of their practitioners. Each organization scheduled a series of practitioner meetings to present on the concepts of data science and machine learning and to share how these tools combined with their data could produce case-specific predictive and recommendation insights. These presentations included sharing the evidence from BCT Partners' precision analytics studies in the fields of juvenile justice and child welfare. Leaders approached these meetings as a pitch to the practitioners to get their buy-in; if practitioners were too resistant, leaders knew the project wouldn't work.

Practitioners had conflicting immediate reactions: they were excited to receive real-time feedback on their clients but concerned that their experience-based decision-making authority would be overridden by algorithms and that these algorithms would be faulty. Specifically, their biggest concern was that they didn't want algorithms "telling them what to do"; some were familiar with and shared the public stories of how algorithms could be biased. BCT's evaluation and data science experts participated in these meetings to present previous study findings and listen to and address practitioner concerns.

In response to these concerns, BCT worked with practitioners to co-develop ground rules for the precision analytics approach:

1. The algorithmic process wouldn't use "black box" algorithms like deep learning that hid what was going on inside
2. The modeling would not use client demographic characteristics that didn't or shouldn't play a role in who gets "selected" for different treatments (e.g., race, gender, etc.)
3. Recommendations would transparently share each case's probability of success, with and without the recommendation (i.e., counterfactual evidence), rather than presenting the recommendation as a "must do" without this evidence of how much it matters
4. Practitioners could meet regularly to discuss findings and lessons, and make changes to what was being measured and how findings should be presented

One key driver for practitioner buy-in was the explicit support of the program leaders and managers that was garnered through the establishment of these ground rules. Ultimately, practitioners bought into pilot testing this work because precision analytics acknowledged and leveraged their truth that one size doesn't fit all by providing more tailored lessons, feedback, and recommendations tuned to the unique needs of each client that they serve.

Figure 3: Minimizing Algorithmic Bias When Applying Precision Analytics to Social Impact Program

Only use transparent machine learning algorithms, where the resulting rules can be seen and interpreted by evaluators and practitioners alike. Since neural network and deep learning algorithms don't adhere to this rule, they are not used for PA. PA relies primarily on different types of decision tree algorithms as they meet this criterion.

Don't use demographic characteristics to find matched comparison groups of similar cases, unless there is valid reason for treating persons with specific demographic characteristics differently. For example, during the PA process of finding matched comparison groups to evaluate whether educational programming can help achieve a living wage for youth aging out of foster care, race should not be included as a variable on which to match, unless there is a reason that a youth of color should be educated differently based on their skin color (which is not the case). Put another way, if skin color shouldn't be a program selection criterion, it shouldn't be used to identify matched comparison groups

Evaluate if all demographic groups equally received what works. The PA process concludes with the evaluation of who within the matched group actually got what worked and succeeded. At this point, it is critically important to analyze these findings to ensure that there are no statistically significant differences as to who got what works, based on demographic characteristics.

Securing Funding

With practitioners on board, the CEOs of each organization independently approached their funders to raise the money needed to build the precision models. The cost of the modeling project as well as the design and implementation of practitioner dashboards/tools and real-time evaluation reports was estimated at about $100,000, not including the cost of staff time at each organization. (Since these projects were conducted, "build" costs have generally declined, although they vary according to the quality and quantity of program data.) The cost of the build included creating an automated technological process that cleaned, prepared, modeled, scored and produced findings. New data could be automatically run through the model, on demand. Once the initial models and tools were built, they would cost between 10% and 25% of the build cost annually for ongoing support and maintenance. Based on up-front cost estimates, each organization believed that they could independently sustain their respective "learning systems" and focused on raising funds for the initial build stage.

Gemma Services approached Scattergood Foundation, a grantmaker in the Philadelphia region that focuses on behavioral health and has an interest in programs that improve practice, measure performance, and guide systems change that support children and families facing mental health problems. Upon learning about the potential application of precision modeling, this funder was interested in partnering around the development of actionable evidence not only for Gemma Services and its team, but also for funders and policymakers concerned with mental health service provision across the City of Philadelphia. Resultantly, Scattergood Foundation

facilitated conversations with Gemma and the City of Philadelphia's Department of Behavioral Health and Intellectual Disabilities Services (DBHIDS).

DBHIDS serves as a managed-care funder of residential psychiatric services for children in the City, providing reimbursement to several contracted child residential psychiatric facilities, including Gemma Services. Representatives from Gemma and Scattergood met with the DBHIDS Commissioner to introduce the project and seek support. The Commissioner was highly interested in supporting this project to test the feasibility of using program administration data to determine what works for different children and their families, including being able to learn about which contracted provider might be the best match for different children, as well as learn more about the "ideal" lengths of stay for different segments of the population; he sought to better understand costs. Scattergood Foundation was the lead funder, and the City of Philadelphia and Gemma Services contributed the remaining dollars needed to support the PA project.

First Place for Youth raised money from their funders in two stages. The first funders – Tipping Point Community and The Trustees' Philanthropy Fund of Fidelity Charitable™ – supported an initial modeling process to test the feasibility of precision modeling. Once this project had validated the proof of concept, the organization raised money from the Annie E. Casey Foundation and The Charles and Helen Schwab Foundation for model refinement and to build practitioner and leader decision dashboards. These funders were interested in this project for four reasons: 1) to provide case-specific decision-making actionable evidence (insights) to practitioners; 2) to automate the evaluation process; 3) to begin to develop a "seed" set of data features (variables) and algorithms that could be scaled to other extended foster care providers for networked learning; and 4) to develop more nuanced and precise findings that could be leveraged for policy and advocacy work with state and federal governments.

Partners

With initial funding in place, BCT Partners worked separately with First Place for Youth and Gemma Services to apply precision analytics to their work.[5] Once the organizations engaged, each one independently had to invite and support four key partners to play collaborative and complementary roles (see Figure 4).

[5] Interestingly, BCT connected with both organizations through conferences and professional networks where the focus of the exchanges was on the use of data science to achieve social impact. These connections were not made through traditional evaluation and research events or networks. The author of this paper, himself a 20+ year evaluator, has found that the two fields of data science and evaluation are in their own siloed networks, and that the motivation to learn about using data for practice is being driven more by program leader interests in learning about predictive analytics and data technologies than an interest in evaluation. In fact, to address this divide, the author of this case study was engaged by the Rockefeller Foundation in collaboration with a leading development evaluator, Michael Bamberger, to write a paper on how data science and evaluation need to come together to better track, monitor and evaluate social impact: Measuring Results and Impact in the Age of Big Data: The Nexus of Evaluation, Analytics, and Digital Technology.

Figure 4: Roles in the Precision Analytics Process

Role	First Place for Youth	Gemma
Evaluator • Experienced with designing and conducting mixed methods program evaluations • Responsible for program and analytic frameworks, including developing program logic model (with the practice expert(s)) and selecting/naming dataset variables that align with the logic model • Decision-maker on how to conduct attribution analysis to reduce bias that often comes with algorithmic data science efforts that focus more on correlational prediction than causation • Directs the data scientist during the training of the machine learning algorithms to ensure the reduction of selection and other forms of bias inherent to the data; with the practice expert, leads all algorithmic decision making, including deciding which variables on which to train the models • Provides ongoing evaluation support to the practice expert and organization	**Internal** Dr. Erika Van Buren, Vice President of Learning, Evaluation and Strategic Impact	**External** BCT Partners
Practice Expert • Educated/experienced in providing clinical services to children, youth and their families; comfortable descriptively analyzing program data • Collaborates with the evaluator during the modeling process, serving as an advisor, consultant and liaison to other staff who need to engage in the work • Validates algorithmic findings, including whether matched comparison groups actually represent populations that need to be treated differently and patterns of "what works" for each group; has veto power over final "what works" model, with algorithms re-trained until findings make practical sense • Leads design and implementation of training and support resources and processes for staff/stakeholders	**Internal** Team of program leaders and managers, led by the program director	**Combination** Former employee social worker/clinician Kate Ryan, MSW as lead, engaging two clinical directors and other practitioners
Data Scientist • Trained in extracting, transforming, and loading data; conducting predictive and prescriptive analytics; data visualization; writing code • Extracts, transforms and loads (ETL) the data for modeling • Provides technical support to the evaluator and practice expert during the precision modeling process • Automates every step in the final model, from ETL to scoring and results production • Designs data tools for practitioners, with practice expert and practitioner input, typically using data visualization tools like Tableau or PowerBI	**Combination** Matt Levy, FPFY Director of Data and Performance Management, and BCT Partners	**External** BCT Partners
Practitioners • Represent the users of the eventual tools and dashboards, and have final say as to the usefulness of the generated insights • Provide critical feedback about the professional support, training and capacity building that would be required for them to use findings for case-specific planning, monitoring and evaluation • Provide input on, feedback on, and validation of precision modeling results and dashboards	**Internal** Team of program leaders and managers, led by the program director	**Internal** Clinical directors and practitioners

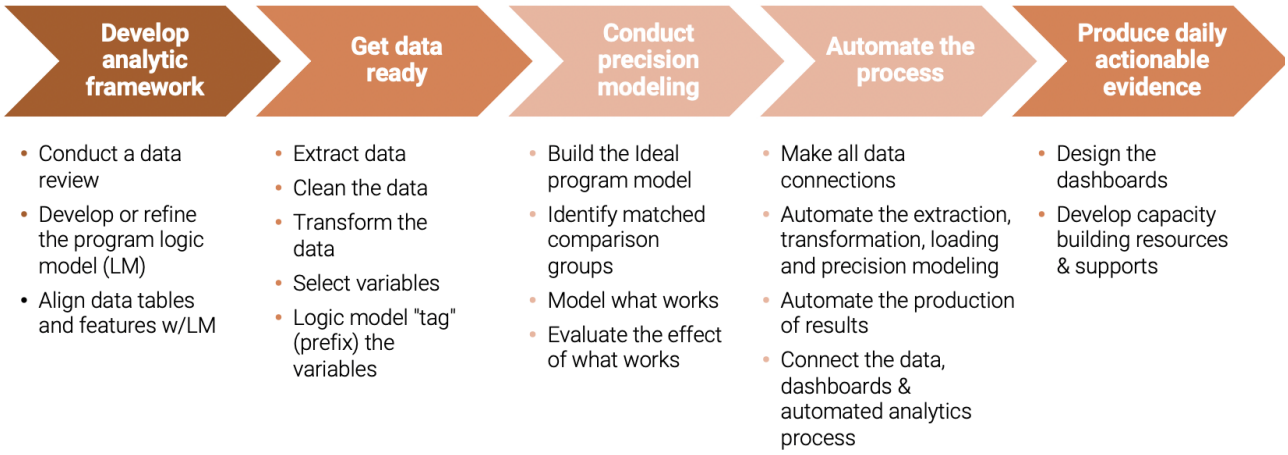
Approach

Confirming Readiness for Precision Analytics

Before beginning the precision modeling process, a data audit is required to evaluate the quality and quantity of data to determine "readiness." Based on prior program administration modeling projects, BCT has determined that the following minimum data qualities are required: (1) a minimum of 250 cases to allow for the disaggregation of cases into matched comparison groups; (2) at least two years of reliable longitudinal intake, program, and outcome data; (3) low levels of missing data; and (4) high levels of variability for key questions/measures. Both First Place for Youth's and Gemma Services' data met these requirements. Additionally, both organizations requested additional preliminary analyses to test the feasibility of generating results during the full precision modeling process. The data audit took ~10 weeks.

After confirming readiness, there were five steps to the precision modeling process for First Place for Youth and Gemma Services (Figure 5). What follows is a description of how each step was carried out, what types of capacity building were required to support the work, and the challenges and remedies that were encountered along the way. The whole process, not including the data audit, took ~36 weeks.

Figure 5: Stages of the Precision Modeling Process



1) Develop the Analytic Framework

The first step was led by each organization's evaluator, who facilitated an iterative process with the practice expert to develop and/or refine the program logic model. In parallel to the logic model development process, the evaluator worked with the data scientists to review and assess the administrative data, including getting to know all of the variables, assessment tools, and dictionaries/codebooks for the database. The last step in the analytic framework process was to align the logic model with high-level data constructs that were well-represented across the administrative data variables. This step required that the evaluator, data scientist, and practice expert come together through a series of screen share

meetings to review data, diagrams, dictionaries, and program descriptions to ensure that the logic model was aligned with the dataset and its variables, and to discuss ideas for transforming data if there were gaps and/or weaknesses in relation to logic model components.

2) Get the Data Ready

This step, which consumes the largest proportion of the total project time, began with the process of extracting the data, based on the analytic framework, from each organization's data system. This step was led by the evaluator, with the lion's share of the labor being conducted by the data scientists. For example, First Place for Youth uses Apricot as their data system. The data scientists had to work with the organization's data manager to query and extract the data, with the support of the evaluator, to ensure the variables were correctly aligned with the logic model. This step also entailed the data scientists cleaning the data, including addressing missing data and variables with too little variance and or skewed responses. They consulted with the evaluator, and sometimes the practice expert, to decide what to do with problem data. Then, the data had to be transformed into the final set of variables that would go into the precision modeling step. These transformations were led by the evaluator, with input and final approval from the practice expert, as well as the technical support of the data scientists. Next, the transformation process created target variables – e.g., using multiple variables to calculate acuity (i.e., severity of a child's current mental health status) for Gemma Services, and determining how First Place for Youth would calculate the target, living wage, when different sites had different costs of living. This step culminated with selecting the final transformed and logic model-aligned variables for the modeling process and creating prefix tags associated with the different logic model components for each of the variable names to help guide the modeling.

3) Conduct Precision Modeling

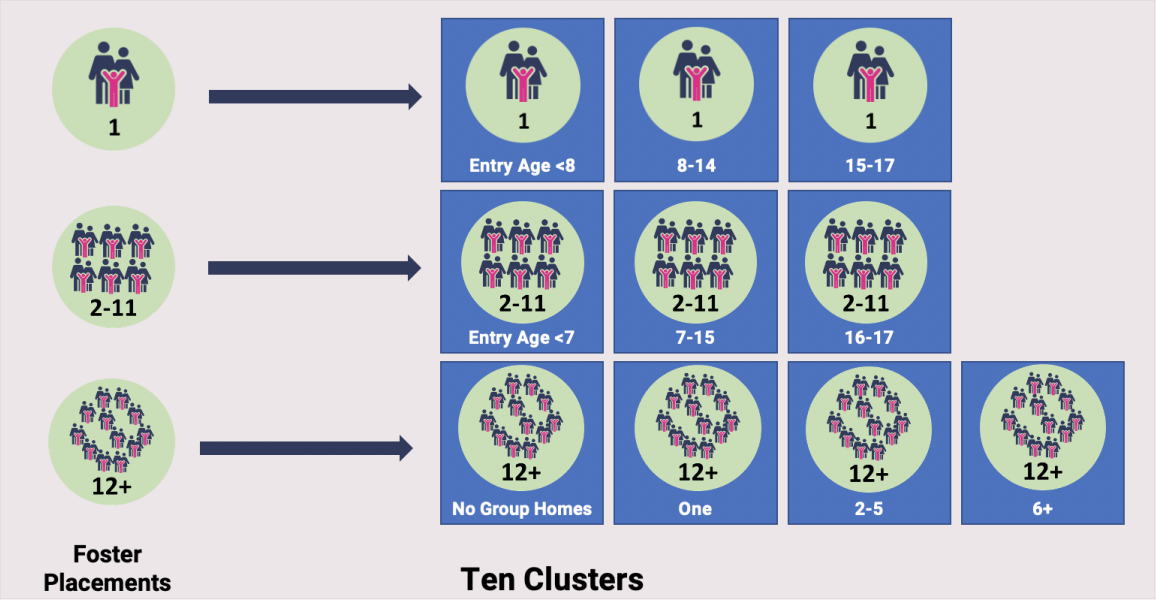
The process of building the precision model entailed the evaluator closely collaborating, through multiple screen share sessions, with the data scientists to conduct a series of modeling steps that found matched comparison groups to reduce selection bias, discovered what works for each group, and evaluated the effect of what works.

The first step in the precision modeling process was to build an "ideal program model." This is a predictive model that uses all of the program dosage, strategy, and goal attainment data to predict the desired outcome. For example, Gemma used their goal attainment data, which represented the tracking of longitudinal changes in thought, behavior, and psychiatric and trauma assessment scores, to predict a child's acuity level at the time of discharge. This first ideal program model was able to identify the goal accomplishments that were most important to reducing a child's behavioral acuity to a level that was associated with a discharge that was much less likely to return to inpatient hospitalization within the subsequent year. This ideal model was about 75% accurate. However, this model had not yet addressed selection bias, as it was built using all children. Gemma Services' practice expert knew that the ideal model couldn't be implemented with all children; one size wouldn't fit all. This goal attainment model

produced a probability for every child as to their likelihood that they would have achieved a low enough acuity to be considered a success, based on their accomplishment of assessment-based goals.

The next step in the quasi-experimental precision analytics process was to identify matched comparison groups based on contextual and baseline intake characteristics that predicted how likely a child was to engage and/or be engaged in the ideal program model.[6] By training machine learning classification algorithms to cluster children into groups based on sharing characteristics that make them equally likely to receive the ideal program model, this step identified matched comparison groups that could be studied and evaluated during the next steps.[7] For example, First Place for Youth's data identified 10 groups based on background characteristics that affected if and how they engage with their ideal program model. (See Figure 6.) Specifically, the number one predictor of if and how a youth will engage in programs is how many foster homes they have been placed in, followed by their age at entry into foster care or how many groups homes they'd been placed in. It makes sense that historical instability and longevity in foster care placements makes some youth less likely to engage in First Place for Youth's programming. As such, it is critical to determine what works for each group separately, thereby controlling for this selection bias.

Figure 6: First Place for Youth's Ten Matched Comparison Groups



[6] This step is analogous to propensity score matching (PSM) statistical procedures used in tens of thousands of health, education, political science, economic, etc. peer-reviewed observational studies to minimize selection bias. However, training machine learning algorithms to identify matched comparison groups mitigates a significant problem that leading social science researchers and statisticians Gary King from Harvard and Richard Nielsen from MIT identified in their 2018 paper, Why Propensity Scores Should Not Be Used for Matching, proving that PSM creates experimental imbalance.

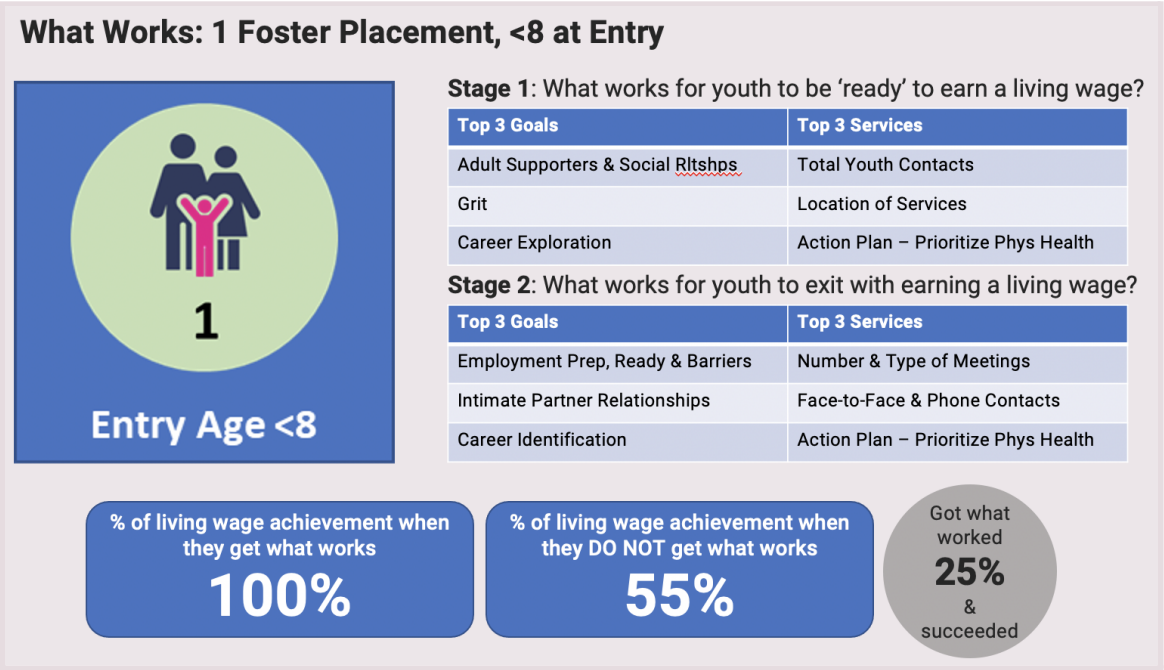
[7] This machine learning process, using simple decision tree algorithms, adheres to King's and Nielsen's recommendation of using fully blocked matching instead of PSM.

The next step in the precision modeling process is to train machine learning algorithms to determine which combination of program interventions, dosages, and/or goal achievements predict the highest likelihood of success for children within each matched comparison group. The evaluator guides the data scientist through an algorithmic training process that identifies the program elements that predict the best outcomes for each matched group. These algorithms are able to produce a ranked and weighted set of program elements that uniquely and in aggregate contribute to achieving the desired outcome. The findings are a group-specific combination of program elements that, when combined, increase the likelihood of a matched cluster of children achieving a positive outcome.

The practice expert is consulted to review and provide validation and explanation for the patterns of what works for each group because of their understanding of both the research and what it takes to implement "promising practices" on a case-by-case basis. If the practice expert judges the pattern to be partially or fully invalid, the evaluator and data scientist make adjustments to which program variables were put into the model and/or in what order, iteratively, until the practice expert validates the explicability and usability of the results.

Below is an example of "what works" for one of First Place for Youth's matched comparison groups. First Place for Youth's what works modeling was broken into two programmatic stages. This diagram shows what works at each program stage for one group of First Place for Youth participants. Each group has a different set of goals and services needs to succeed (achieve a living wage).

Figure 7: "What Works" for One of First Place for Youth's Groups at Two Programmatic Stages



The final step in the precision modeling process is to inferentially evaluate the effect that the group-specific program model had – in the past – on a matched group of children when some received what works and some did not. The analytic process included conducting inferential statistical tests (e.g., t-tests, ANOVAs) to determine if those within a specific group who received what works achieved a significantly better outcome score than those in the same group who did not. Effect sizes were also calculated. For example, for the same group of First Place for Youth beneficiaries, there was a statistically significant difference with a large effect size for achieving a living wage when members of this population received what works, compared to those that did not. In fact, as shown in Figure 7, when youth received their group-specific set of services and achieved their goals they were almost twice as likely to achieve a living wage. Lastly, the precision modeling can tag every case with one of the following four outcomes:

1) Attributable success: Got what works and succeeded
2) Independent success: Didn't get what works, but succeeded anyway
3) Unmet needs: Didn't get what works and didn't succeed
4) Unknown needs: Got what works but didn't succeed

As Figure 7 shows, the program has an attributable success rate of 25%.

4) Automate the Analytic Process

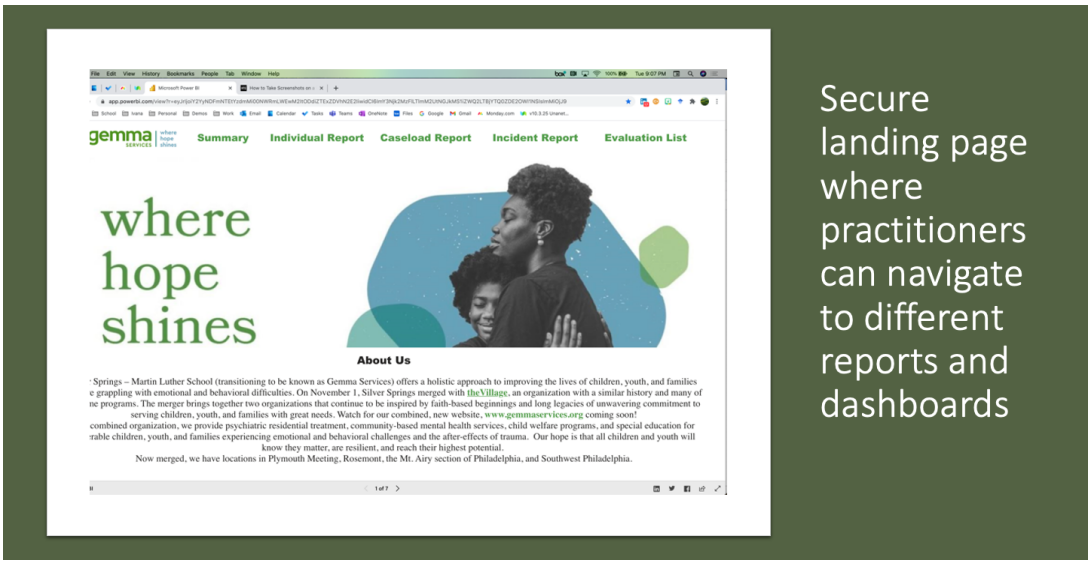
The fourth step in the process is to engage the data scientists in automating the analytics workflow such that data extraction, transformation, and loading, precision scoring, and results generation all happen at least once a day without requiring any human to run. The automation includes making all data connections to an organization's data system, using an API or other data connection method. The automation also connects the final, scored results to the dashboarding tool/software (e.g., Tableau or PowerBI). This step often requires engaging an organization's information technology staff and/or database manager. This step also requires determining how to keep the data secure, confidential and compliant with relevant data privacy standards such as HIPAA throughout the entire data transfer process. The workflow and connections can be made behind the client's firewall and security measures, or can be encrypted in transit at each connection stage and kept secure in a cloud-based architecture; there are a number of secure options that allow real-time, on-demand utilization. While this process may sound technical, time consuming and expensive, it is not. The process for both First Place for Youth and Gemma Services required a few hours of setup and two to three 30-minute meetings between the data scientists and each organization's IT staff. The cost of ongoing support and maintenance (10% to 25% of the build cost annually) includes resourcing the automation and security procedures. This step can also be completely outsourced for organizations that do not have the technical staff capacity to assist with or support the process of making secure connections.

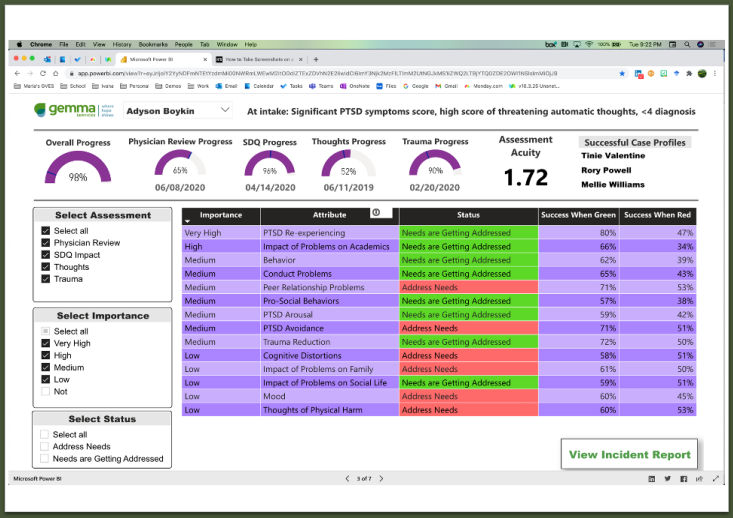
5) Produce Daily Actionable Evidence

This final step is to design and implement a suite of dashboards for an organization's practitioners, program managers, and leaders to receive on-demand insights – actionable evidence – to be used for case-specific, programmatic, and organizational decision making. This step required data scientists with design and visualization training and experience. Most organizations now begin with templates created by BCT that use software programs like Tableau and PowerBI to generate dashboards, reports, and visualizations for practitioners.

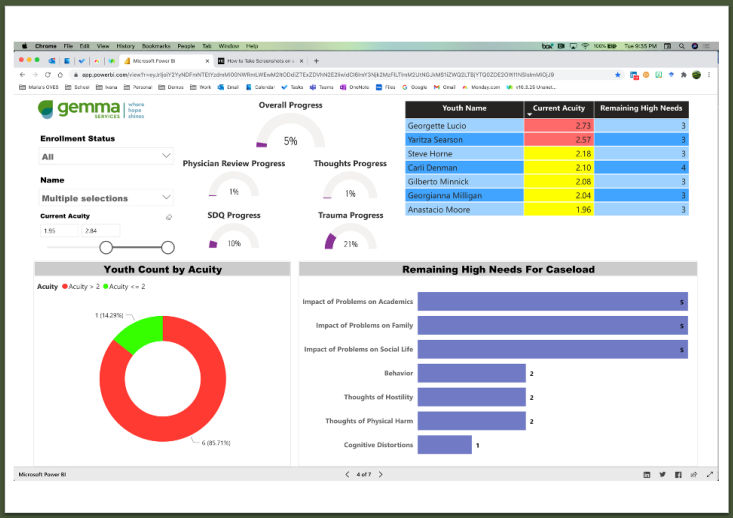
What follows are screenshots from Gemma Services' "learning system."

Figure 8: Screenshots from Gemma Services' Learning System (Case Names Changed)

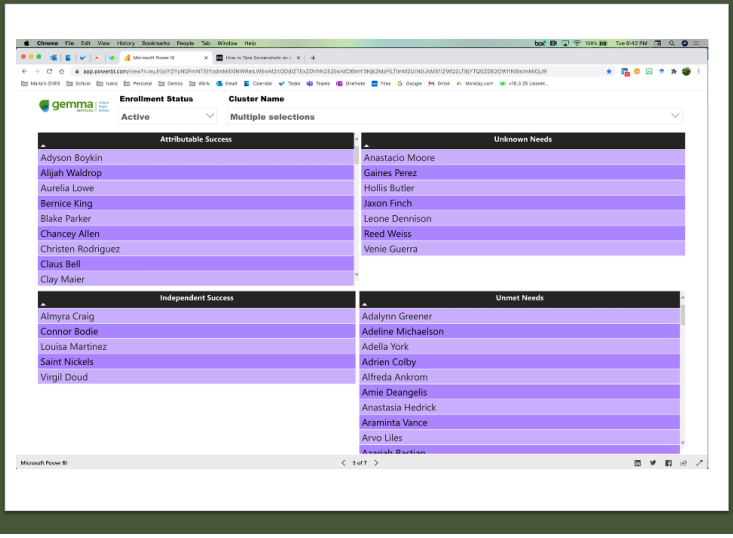




A dynamic case-level dashboard presenting up-to-date progress indicators; individual needs (attributes) with the child's current status and counterfactual evidence of success based on similarly matched youth from the past; effect size descriptions that serve to rank order the needs; and names of similar children with whom the organization has succeeded in the past, to qualitatively investigate the story of what worked.



A caseload dashboard where practitioners can dynamically investigate how their assigned children are doing, including which high need areas are the top priorities for most of their cases.



A dashboard that allows practitioners to identify specific cases by their current outcome level. This tool is used to qualitatively investigate these cases to compare and contrast what might be going on, including that which the data may not be capturing.

In support of these technology solutions, each organization also developed and is evolving a set of practice resources, documents, training materials and tools that complement their respective learning systems. For example, First Place for Youth's program leadership engaged in a series of steps to develop a set of resources and guides for all practitioners that they call their "Youth Roadmap." In fact, they are labeling their precision learning system the "Youth Roadmap Tool" (YRT). Their YRT will include links within the dashboards to connect to specific sections and resources within the Youth Roadmap toolset. Gemma Services has also developed a set of documents, resources, and guides that help support the insights deriving from their learning system, which is called the "Outcome Generator." They too are integrating links within their dashboards to make specific support materials available on demand. Additionally, the dashboards make use of mouse-over effects that explicate key concepts, define terms, and provide quick guidance on tool utilization.

The five-step process described above has been streamlined to be completed in 14 to 20 weeks, at which time up-to-date actionable evidence is available via a suite of dashboards. Because the process has been codified and leverages big data analytics and technologies that have come down in price significantly, it will continue to be a cost-effective, sustainable and timely way to deliver real-time actionable evidence to practitioners and their clients.

It is important to note that external and internal funding, policy, population, programming and practice changes will occur periodically, requiring the models to be retrained. To accommodate these changes, both organizations plan on retraining their models every two years. This process will require 4-6 weeks to rerun the training using the automated data preparation, precision model, and results production process; revise the dashboards; and update training resources and guides.

Challenges and Responses

Challenges occurred at each step in the precision modeling process for First Place for Youth and Gemma Services. Below are the five top challenges and how they were addressed:

1. **Low-quality program intervention data.** Many program datasets do not capture enough structured data on the qualitative experience to provide rich explanatory insights. The program monitoring variables in organizations' administrative datasets are often timestamped attendance data by type of service, activity, or event. These data don't typically offer much, if any, information about the modalities (e.g., cognitive behavior therapy), methods (e.g., 1:1 therapeutic session), and/or breakdowns of how time was spent. The best solution for this problem was to make data improvements that would benefit future precision modeling updates by integrating stronger questions into the administrative data collection system. For example, Gemma Services' practice expert and evaluator needed to address having too little service dosage and family engagement data to adequately represent these components of their logic model. Their solution was to develop a family engagement survey that would start to be administered and modeled 12 to 18 months after the initial model was up and running. While this was a limitation of the

initial model, one advantage of the precision analytics approach is that the model can be updated as richer information becomes available.

2. **The need for capacity building around the unique set of methods, tools, and approaches of each team member's area of expertise.** Those in each role – practice expert, data scientist, and evaluator – brought valuable expertise to the project but lacked experience in other areas. This was resolved through informal and formal capacity building, including co-learning by observing and doing, and intentionally setting aside time for formal orientation to and high-level training on each role's area of expertise, whether methodological, technical, and/or practical. For example, data scientists required some capacity building on logic models and their use and importance for framing the subsequent steps of getting the data ready for conducting precision modeling; practice experts required training on the principles of precision modeling; evaluators needed training on machine learning and how it is different from statistical modeling; and data scientists and evaluators needed training on the provision of services related to extended foster care for youth aging out of the child welfare system (First Place for Youth) or residential psychiatric services for children (Gemma Services). Additionally, the team learned by doing, using an analytics software, KNIME, that was visually based and didn't require any coding to use. Using KNIME, team members could understand what was going on in the modeling process, and even contribute to improving the analytics process and workflow as the process unfolded. The non-data scientists did require some orientation to the workflow's visual components, but KNIME made the process much easier to grasp than it ever would have been with code.

3. **A lack of understanding of causal modeling and how selection bias could be mitigated when using administrative data for evaluation.** A significant challenge with gaining buy-in to the precision modeling process was educating the practice experts and practitioners on what selection bias is and why it needs to be reduced. The evaluators used an approach to explaining causal modeling through mitigating selection bias that ultimately helped everyone grow in confidence that the process would result not just in correlational predictions but in valid causal inferences (albeit less certain inferences than with a randomized controlled trial). The process of unpacking selection bias began by asking, "If some other similar program selected youth who weren't as troubled as your program, would it be fair to expect your program to achieve the same level of success?" Practitioners would respond no, stating that it would take more and different types of programming because their youth were facing more and different challenges. We explained that this element is also at work within their programs, in that not all youth are the same and different groups need to be "selected" for different types and levels of services to succeed. This made intuitive and experiential sense to practitioners, which bridged the technical gap of understanding why we need to reduce "selection bias" by finding and evaluating matched comparison groups of similar youth.

4. **Achieving practitioner buy-in/confidence in the approach when analytic causal patterns didn't reflect their experience.** At times, although analytic confidence in a relationship within the data was high, the practice expert expressed doubt in the modeling results. In these situations, the evaluator's role was to engage the practice expert in explaining why they thought the pattern appeared, rather than just declaring that it couldn't be true. Often, a discussion that encouraged the practice expert to generate an explanation for why the pattern could be true – a conversation imbued with the research philosophy of assuming the opposite of one's hypothesis is true (i.e., the null hypothesis) – resulted in the practice expert coming to a more pattern-aligned conclusion than their original judgement. Other times, the practitioner's explanation led to the remodeling of the data that resulted in a more valid result. This learning process, with deference to the practitioner, not only ensured confidence in the final results, it also ensured face validity for the practitioners who carried their support of the final tool to their peers who had not been a part of the process.

5. **Avoiding the perception that this was about performance assessment of practitioners.** Most practitioners embraced the opportunity to have data-driven, real-time feedback to guide their work and make course corrections. At the same time, practice experts and program managers quickly realized that the dashboards could be seen as a performance assessment tool for practitioners and that this could be problematic. For example, if a practitioner's cases weren't making progress (such as on the progress gauges shown in screen shots in Figure 8) or, worse, were regressing, practitioners were worried about being judged as ineffective before there was an understanding of what unmeasured factors might be contributing. The managers addressed this philosophically by sharing that the focus of the tools was on each child or youth and ensuring their progress, not on job performance. More practically, practice experts worked with managers to develop a set of practice tools, guides, and resources specifically addressing each recommendation that could show up on a child's or youth's dashboard, so that practitioners could easily access and review how to implement what was being suggested. These included creating links to practice guides and resources right within the dashboards. Additionally, there was a page created that identified past children or youth who achieved the desired outcomes so practitioners could qualitatively investigate these cases by pulling their information and talking with the practitioners who had previously worked with these successful cases. Each of the organizations' practice experts and/or program managers regularly meets with practitioners to share lessons, deliberate on cases, and conduct deeper inquiries on cases that require more investigation.

Results

First Place for Youth and Gemma Services have begun implementing their dashboards and practitioners are now using actionable evidence, in real time. The dashboards and tools resulting from precision modeling are designed to be actionable not just for front-line practitioners but also for program developers and managers, organizational leaders, and system change agents, including policy makers and funders. Additionally, the goal of these

tools is to strengthen program delivery in a tailored way that will improve outcomes. The evaluation tools provide the insights, while resource guides and the learning and professional support of peers and managers help practitioners take action to increase the proportion of those who achieve the desired outcomes.

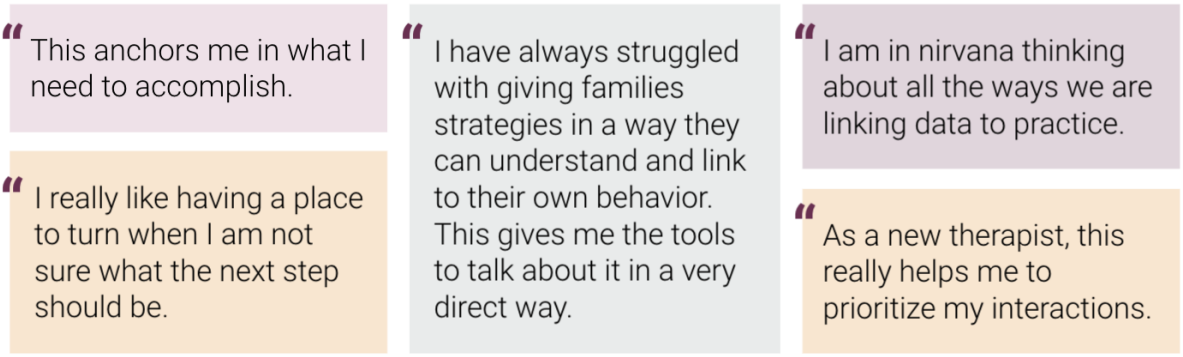
The precision analytics process and associated tools led to a culture shift in how practitioners at First Place for Youth and Gemma Services do their work and how performance is evaluated. Prior to the development of the learning and evaluation dashboards, progress and results of practitioner-led case-specific decisions were not available to practitioners (and were certainly not available in real time, updated on a daily basis). It is now possible to measurably view the progress, performance, and outcomes of each child or youth in a practitioner's caseload, as well as the performance of their whole caseload, with updates on a daily basis.

Organizations are monitoring and tracking the utilization rates for their tools. Currently, each organization's program leaders, managers, and front-line practitioners are accessing and using their respective tools between 300 and 500 times per month.

Some of the early results of having immediate and up-to-date actionable evidence include:

- **Practitioners use case-specific dashboard evidence for case engagement.** For example, clinicians working with children in residential milieu at Gemma Services are using the individual dashboard to understand more about the needs of the children they are working with on a daily basis.

Figure 9: Quotes from Gemma Services Practitioners



- **Program directors use cluster evaluation information, including the names of children/youth at different current outcome levels, for case planning.** Program directors at both organizations are finding it helpful to be able to identify all of the cases at different outcome levels – i.e., attributable success, independent success, unknown needs, and unmet needs. This allows directors to meet with front-line staff to investigate why different youth are in different places, examine what the data are beginning to indicate, and, most important, qualitatively examine the root causes for a child's outcome status. Program leaders find that these inquiries, made possible by

having up-to-date evidence, are strengthening case plans. Real-time feedback also allows for rapid plan changes.

- **The precision tools are motivating program directors and front-line practitioners to gather more data**. In the past, both organizations faced challenges with getting assessments completed on a regular basis. Now, with dashboards that update in real time, including indicating the date when the last assessment was conducted and/or data was entered, practitioners and program managers are visually cued to gather data more frequently and/or on time. They are motivated by their desire to monitor the progress of the youth with whom they are working. They want to ensure that case-level evidence presented in the dashboard is up to date, which often requires updating a key assessment.

- **The precision tools are engaging practitioners to help improve the data.** As more program directors, managers, and practitioners make deeper qualitative inquiries into their cases and what is going on, they are beginning to realize that they need additional data points to test hypotheses that cannot be answered by the current data. For example, the Gemma Services practice expert learned that program directors felt strongly that parent engagement was a critically important variable that wasn't being captured in the data. They hypothesized that the quantity and quality of parent engagement, with both the child and the clinical staff, were key determinants of achieving many of the goals needed to reduce acuity. So, the practice expert worked with the program directors to design a set of questions that will be tested and modeled in the near future for use on an ongoing basis. There may be times in which new questions need to be developed for specific matched groups due to their unique needs or circumstances. Milieu staff at Gemma Services wanted to improve the data pertaining to the role that daily milieu behavior plays in a child's progress. In response, the practice expert and program directors designed a set of structured questions that would be completed when milieu workers entered their shift notes on a daily basis. Their hope is that this data will enhance the models and evidence even further.

- **Precision modeling findings are being leveraged for policy change.** First Place for Youth's Director of Public Policy, Vice President of Learning, Evaluation and Strategic Impact, and academic research partner have used the findings from their precision modeling process to write a policy brief, Raising the Bar: Building System- and Provider-Level Evidence to Drive Equitable Education and Employment Outcomes for Youth in Extended Foster Care. The purpose of this paper is to encourage the State of California and federal policymakers to scale First Place for Youth's extended foster care model, in conjunction with their Youth Roadmap Tool learning system.

What's Next?

Creating Learning Hubs

Both First Place for Youth and Gemma Services are raising funds to scale their learning systems to create a learning network of similar organizations everywhere, all learning together by adopting and/or building their own learning systems. In fact, Scattergood Foundation plans to fund Gemma Services to scale their models to other residential mental health providers throughout the region, state, and country. The vision for both organizations is to become the collective learning hubs for similar programs in communities throughout the United States. First Place for Youth is beginning to plug new affiliate programs throughout the U.S. into their Youth Roadmap Tool. If a similar program doesn't have enough data, they could, for example, get started by adopting First Place for Youth's extended foster care question sets, algorithms, and dashboards that use standardized background and context data to identify different groups of youth, as well as adopt the validated assessment tools that provide insights about the unique needs of each group of youth. This will provide preliminary insights about the tailored goals needed and commensurate progress made for each youth. Then, once 12 to 18 months of program delivery data have been added, the clustering and goals tool can be enhanced to include program-specific recommendations. If they have enough of their own program administration data, they could build their own precision models. At this point, the build cost is affordable to most larger organizations, and the ongoing support and maintenance cost is sustainable for medium to large organizations. First Place for Youth and Gemma Services, in their respective program areas – i.e., extended foster care and child mental health services – would serve as learning hubs, helping to support the capacity building needs of network members and, most important, leveraging system-wide lessons, insights, practices and studies to effect practice and policy change, locally, regionally and nationwide. This is their vision, and they have funders currently interested in supporting their efforts.

Engaging Intended Beneficiaries in the Process

Currently, practitioners are the users of the precision analytics tools on behalf of their beneficiaries. Both First Place for Youth and Gemma wanted to focus on the practitioner for their first precision projects to better understand what the process and engagement actually entailed. But both organizations also voiced a goal of eventually engaging beneficiaries (youth, in the case of First Place for Youth; and children and their families for Gemma Services) in directly using the precision learning tools so they can actively participate in interpreting outputs and planning what to prioritize to improve the likelihood of success. Additionally, both organizations would like to engage beneficiaries in capturing the information they think is needed to better understand what is working and what is not. This includes engaging beneficiaries in the design of new questions and tools for capturing their perspective ("voice"). Gemma and First Place for Youth have plans for beneficiary engagement once they have had time to study the early implementation by practitioners.

Taking on Field Challenges

There are some key challenges ahead. For one, evaluators and practice experts need to be educated about the use of machine learning algorithms and big data analytics for social science research and evaluation. They need this knowledge to develop an understanding of how to collaborate with data scientists to build these tools. Enough precision modeling projects have been done over the past six years to codify processes and generate strategies to educate the sector, including producing peer-reviewed and professional publications and methodological and findings presentations. There are government agencies that have tested the feasibility of and/or are using precision modeling in their work (e.g., the National Science Foundation's Evaluation Office, the Department of Health and Human Services' Administration for Community Living, and Northwest Pennsylvania's Job Connect, a state-funded workforce board). The American Evaluation Association's President-Elect, Veronica Olazabal, who is also Director of Evaluation at The Rockefeller Foundation, has worked with the author of this case study and other prominent evaluators such as Michael Bamberger, a top international development evaluator, to share how big data and big data analytics is changing the field of impact measurement. Presentations have been delivered to large audiences of evaluators and researchers, like AEA's and the European Evaluation Society's members. However, more dissemination is needed, and more engagement in learning about data science techniques is needed in the social science and evaluation communities.

Another key challenge is having enough data to scale this type of work. Many nonprofit providers don't have the 250+ cases of longitudinal case-level data to get started. However, as the cost of program administration systems and the number of vendors grow, there are many more organizations that have and/or are in the process of setting up and implementing robust program administration data systems. So, there are many more organizations that are getting and/or are ready. The learning networks described above may also provide on-ramps for organizations not yet ready to create their own modeling and tools.

A third challenge is ensuring that all identifying data are protected and secure. Technologies are now in place that, whether within organizations or in the cloud, protect the identity of cases in datasets. Organizations that were part of a learning network wouldn't have to share data, but could keep it secure behind their own firewall, or behind the secure firewalls of HIPAA-compliant cloud platforms like Amazon AWS. The learning hubs could leverage aggregate findings. If there is a desire to share data, there are efforts of organizations like Brighthive that create Data Trusts, which both technologically and through policy agreements create shared data architectures and processes that protect information.

Evaluating the Learning Systems

As interest in the development and use of precision analytics tools grows and organizations like First Place for Youth and Gemma Services deploy them, it will be important to evaluate if and how practitioner use of the tools is improving beneficiary outcomes. Subjectively, organizations can and should conduct pre-tool deployment and post-tool deployment evaluations of each of the matched comparison groups' outcomes to inferentially determine if

outcomes improved for each group after implementing the use of the tools. Because data system use is being tracked within the administrative data systems, including how much each practitioner is using the system, this longitudinal evaluation could include controlling for utilization rates. There may even be some practitioners who don't use the system at all, providing naturally occurring experiments on which to focus the outcomes evaluation. Objectively, it will be important for external evaluators to assess the precision analytics tool's use on beneficiary outcomes through conducting more rigorous evaluations like randomized controlled trials. Additionally, outside evaluators could gather more qualitative data on if and how these tools are affecting practitioner behavior.

Alignment with Actionable Evidence Principles

Principle	In This Case...
Centers on Community Needs and Voices *Addresses the context, perspectives, priorities and assets of students and families, along with the challenges they face*	First Place for Youth and Gemma used data analytics to disaggregate overall populations into subgroups of children and youth based on unique backgrounds, histories and context. This allowed for building models and practitioner tools that weren't "one size fits all," but instead produced recommendations tailored to each subgroup's needs, maximizing each case's probability of success. There are future plans to develop strategies for engaging youth and/or families directly in precision analytics work.
Prioritizes Practitioner Learning and Decision-making *Answers questions that are highly relevant to policy and practice, and that help practitioners prioritize decisions in service of students and families*	The data analytics approach and solution address practitioners' key challenge of trying to implement evidence-based practices with fidelity when each case is unique. By building models and recommendation tools based on what historically maximized success for similarly matched cases, front-line practitioners receive timely guidance tailored for the child or youth they are treating.
Enables Timely Improvements *Allows practitioners to make evidence-informed decisions in a timely manner*	The data-driven insights provided to each organization's practitioners are automated and updated on a daily basis. Practitioners, program directors, managers, and organizational leaders all receive quasi-experimental outcome findings, also updated daily. Both organizations are using the real-time insights to more deeply investigate cases qualitatively and refine programming, as well as to measure what they are doing and the results.
Credible and Transparent *Uses high-quality data and analysis, aligning methods with practitioner questions, timeline and context*	Well-established techniques for observational data are applied, such as reduction of selection bias through rigorous matching techniques (with machine learning applied to this task). Additionally, Gemma and First Place are intentionally not using "black box" algorithms, which ensures transparency and accountability.

Responsive to Operational Context of Practitioners *Reflects the context in which practitioners operate, including organizational settings, relationships and resources, and political and policy environment*	The precision modeling approach addresses the need for practitioners to receive insights that are contextually nuanced and relevant to the individual child or youth with whom they are working. Additionally, practitioners receive a list of options for a case with the associated evidence about how much of a difference a particular program element made with similar prior cases. This allows practitioners to select program elements that are most helpful and feasible given current resource realities.
Accessible and User-Centered *Clearly communicates research design, analysis, and findings to facilitate practitioner understanding and use*	Practitioners are engaged throughout the precision analytics process to ensure shared understanding and buy-in. Practitioner dashboards are designed with leadership, input, and feedback from practitioners, using a human-centered design approach. Daily updating of data and dashboards, including progress indicators, provides practitioners with incentives for data entry and use.
Builds Practitioner Capacity for R&D *Provides practitioners with data, products, tools and trainings to own and advance their evidence agenda*	Practitioners are proactively generating new hypotheses, qualitatively investigating cases, and creating new questions/metrics to test. The real-time updated feedback they receive on every case is prompting these R&D efforts. Program managers and directors are using the insights to spend more time learning with and from the practitioners they lead and manage.
Attends to Systemic and Structural Conditions *Considers systems, policies, practices, cultural norms, and community conditions that drive inequity, including those related to poverty and racism*	To identify potential systemic bias, the predictive analytics process examines whether all demographic groups have equal access to "what works" programmatically. To avoid reproducing any existing bias, Gemma and First Place do not match cases based on sociocultural factors (e.g., race) that shouldn't be a factor in why a child or youth gets "selected" for different program options.

Reflections and Conclusion

First Place and Gemma shared a need for a particular type of actionable evidence: they wanted to leverage insights from past program participants to help front-line staff better serve each youth in front of them.

In this case, producing actionable evidence began with looking beyond traditional methodologies. First Place and Gemma understood that their existing data systems, while useful for other purposes, weren't generating the information that practitioners most needed — and a traditional impact evaluation such as a randomized controlled trial wouldn't either. BCT Partners' precision analytics approach offered an alternative. BCT worked with each organization to confirm that their data would support precision analytics (and to address initial gaps), to ensure that the results of the modeling were not just mathematically but

practically and ethically sound, and to design accessible dashboards for staff and managers. By working with BCT Partners, which offered both technical expertise and a commitment to meaningfully engaging practitioners throughout the process, and with the support of funders who saw value in this creative approach, both organizations developed learning systems that generated the actionable insights they needed.

Resources and Further Reading

The Book of Why: The New Science of Cause and Effect

Measuring Results and Impact in the Age of Big Data: The Nexus of Evaluation, Analytics, and Digital Technology

A New Way to Use Data: Precision Care for Better Outcomes in Psychiatric Residential Treatment for Children

Predictive and Prescriptive Analytics, Machine Learning and Child Welfare Risk Assessment: The Broward County Experience

Raising the Bar: Building System- and Provider-Level Evidence to Drive Equitable Education and Employment Outcomes for Youth in Extended Foster Care

Using Predictive Analytics and Machine Learning to Improve the Accuracy and Performance of Juvenile Justice Risk Assessment Instruments

Why Propensity Scores Should Not Be Used for Matching

About the Author

Peter York is Principal and Chief Data Scientist at BCT Partners. He has over 20 years of experience as a consultant and researcher in the evaluation and nonprofit fields, and as a national spokesperson for social impact and impact measurement issues. He has spent the last ten years developing analytic techniques that leverage machine learning algorithms and big data to create predictive, prescriptive, and rigorous evaluation (causal) models and tools for social change agents in many fields, including nonprofit capacity building, child welfare, juvenile justice, workforce development, adult justice, adult protective services, and child mental health. Related publications include: *Predictive and Prescriptive Analytics, Machine Learning and Child Welfare Risk Assessment: The Broward County Experience*; and *Measuring Results and Impact in the Age of Big Data: the Nexus of Evaluation, Analytics, and Digital Technology*. He leads the development of BCT's national Equitable Impact Platform (EquIP), a big data platform combining IRS 990, American Community Survey, and geospatial data to assess, evaluate, and study the nonprofit sector's contribution to equitable community improvement. In addition, he has spent the past 20+ years designing and leading numerous research and evaluation studies with private philanthropies, corporations, nonprofit organizations, and government agencies. He has authored book chapters, academic and professional articles, and a book on the topic of evaluation – *Funder's Guide to Evaluation: Leveraging Evaluation to Improve Nonprofit Effectiveness*. He is a popular speaker on evaluation, capacity building, and data science/analytics, presenting regularly at professional conferences.

Author's Acknowledgments

This paper and the work it represents would not have been possible without the leadership and contributions of Dr. Erika Van Buren, Vice President, Learning, Evaluation and Strategic Impact, and Matt Levy, Director of Data and Performance Management, at First Place for Youth; and Dr. Kristen Gay, President and CEO, and Kate Ryan, Consultant, at Gemma Services. These actionable evidence projects were made possible through the generous funding support provided from Scattergood Foundation; the City of Philadelphia's Department of Behavioral Health and Intellectual Disability Services; Tipping Point Communities; The Trustees' Philanthropy Fund of Fidelity Charitable™; Annie E. Casey Foundation; and the Charles and Helen Schwab Foundation. Lastly, and most importantly, the projects shared in this paper would not have succeeded without the game-changing input, feedback, guidance and support provided by both organizations' program leaders, managers, and front-line practitioners.

Case Keywords

K-12 | post-secondary | program improvement | administrative data | data science | machine learning algorithms | dashboard | nonprofit service provider | foundation | external evaluator | internal evaluator | employment outcomes | mental and behavioral health outcomes