Residential Treatment for Children & Youth

ISSN: 0886-571X (Print) 1541-0358 (Online) Journal homepage: www.tandfonline.com/journals/wrtc20

Participatory Evaluation of Practitioner Utilization of an AI-Based Decision Support Tool in a Psychiatric Residential Treatment Facility

Margaret L. McGladrey, Kate Ryan, Corinne Andriola, Carly Meyers & Joanne Trinkle

To cite this article: Margaret L. McGladrey, Kate Ryan, Corinne Andriola, Carly Meyers & Joanne Trinkle (28 Jul 2025): Participatory Evaluation of Practitioner Utilization of an AI-Based Decision Support Tool in a Psychiatric Residential Treatment Facility, Residential Treatment for Children & Youth, DOI: 10.1080/0886571X.2025.2538855

To link to this article: https://doi.org/10.1080/0886571X.2025.2538855

Published online: 28 Jul 2025.

RESIDENTIAL TREATMENT FOR CHILDREN & YOUTH
https://doi.org/10.1080/0886571X.2025.2538855

Routledge
Taylor & Francis Group

Participatory Evaluation of Practitioner Utilization of an AI-Based Decision Support Tool in a Psychiatric Residential Treatment Facility

Margaret L. McGladrey[a], Kate Ryan[b], Corinne Andriola[c], Carly Meyers[a], and Joanne Trinkle[c]

[a]Department of Health Management and Policy, University of Kentucky College of Public Health, Lexington, Kentucky, USA; [b]Quality Assurance, Gemma Services, Philadelphia, Pennsylvania, USA; [c]Center for Innovation in Population Health, University of Kentucky College of Public Health, Lexington, Kentucky, USA

ABSTRACT

Child-serving agencies are increasingly adopting decision support tools informed by sophisticated computational methods and Artificial Intelligence (AI). However, research on the practical application and usage of these tools in child and family services remains limited, despite their potential to perpetuate systemic biases. This paper provides a case study of the implementation of an AI-based decision support tool at a Psychiatric Residential Treatment Facility. Through Ripple Effects Mapping, a participatory evaluation method, the authors conducted focus groups with 31 staff members across different roles. The results revealed that practitioners used the tool to track longitudinal progress; enhance stakeholder communication; and provide person-centered care, such as better matching staff to youth. Implementation barriers included concerns about data accuracy, new staff training needs, and challenges integrating the tool into workflow. The study identified strategies for improvement, including incorporating the tool into daily practices across roles, recognizing front-line staff use of the tool, and investing in training. While AI-based decision support tools must be carefully designed and implemented to avoid perpetuating systemic discrimination, the findings suggest that they can help providers offer appropriate and person-centered care.

PRACTICAL IMPLICATIONS

- Providing coaching and intervention-specific practice integration with supervisors is essential to successful implementation of decision supports.
- Ongoing technical assistance on using decision support tools should inform continuous quality improvement by supervisors and administrators.
- Decision support tools' suggested interventions must be based on diverse input and adaptable to program and field changes.

KEYWORDS

Artificial intelligence; decision support tools; psychiatric residential treatment; participatory evaluation; precision analytics; Ripple Effects Mapping

CONTACT Margaret L. McGladrey margaret.mcgladrey@uky.edu Department of Health Management and Policy, University of Kentucky College of Public Health, 111 Washington Ave, Suite 102, Lexington, KY, 40508, USA

Introduction

To align supports and treatment with family and child/youth needs and strengths, statistical modeling, including Artificial Intelligence (AI)-based precision analytics and machine learning, is increasingly being deployed to build decision support tools for child-serving agencies. These decision support tools recommend data-driven strategies to inform provider decisions with the goal of improving child/youth outcomes (Hall et al., 2023). Studies evaluating decision support tools have occurred in diverse settings, including residential treatment centers, behavioral and mental health rehabilitation organizations, schools, child welfare centers, therapeutic foster care centers, and juvenile justice (Chatpreecha & Usanavasin, 2023; Chor et al., 2023; Clausen et al., 2023; McIntosh et al., 2010; Stewart et al., 2020; Theall et al., 2022; Troy et al., 2021; Trudeau et al., 2023).

This literature underscores the importance of clinician, practitioner, child/youth, and parental engagement in the development and implementation of decision support tools (Barnett et al., 2018; Clausen et al., 2023; Elkington et al., 2023). However, there is limited literature investigating how practitioners use decision support tools after their development, despite their goal of informing decision-making (Barnett et al., 2018; Clausen et al., 2023; Elkington et al., 2023; Shimshock et al., 2022; Van Der Put et al., 2017). Each of these five studies concluded that although decision support tools can be valuable for improving child/youth outcomes, it is critical to understand how practitioners use them to mitigate bias.

Evaluation of practitioner usage of decision support tools that employ machine learning and predictive analytics is particularly important, given the potential for AI-based tools to replicate social biases and discriminatory risk assessment practices in healthcare, the workforce, education, child welfare, and the criminal legal system (Bauer & Lizotte, 2021; Buolamwini, 2023; Caliskan et al., 2017; Eubanks, 2018; O'Neil, 2016; York, 2021). Dr. Joy Buolamwini (2023) describes this systemic potential for harmful bias in AI-based tools as the "coded gaze" of encoded discrimination and exclusion. It is essential to evaluate how practitioners apply AI-based tools in light of endemic racial disproportionality in the child welfare system (Dettlaff & Boyd, 2022; LaBrenz et al., 2023).

The authors believe this qualitative participatory evaluation is the first effort to assess how residential treatment facility staff are using an AI-based decision support tool and how it might be impacting practice and program outcomes. We offer a case study of the implementation of the Insight Generator, an AI-based decision support tool, at Gemma Services, a Psychiatric Residential Treatment Facility (PRTF) in the northeastern U.S.

Methods and Materials

Context

Setting: Gemma Services

Gemma Services provides community-based behavioral health, prevention, and child welfare programs, an approved private school, and a Psychiatric Residential Treatment Facility (PRTF) through three locations in the Greater Philadelphia region and via Telehealth. With approximately 500 employees, Gemma Services accompanies more than 3,000 children/youth and families each year.

Insight Generator Design

Insight Generator is an AI-based decision support tool that uses Gemma Services' own data and precision analytics with the goal of improving PRTF Program outcomes and reducing costs. In 2018, the Gemma Services CEO partnered with data scientist Peter York to write a white paper detailing the use of precision analytics in behavioral health and their process for developing the Insight Generator (Gay & York, 2018). The Insight Generator extracts historical program assessment and clinical data and then applies precision analytics to identify relationships in service utilization data, aiming to improve post-discharge outcomes through clinical consistency and community-based care continuity (York, 2021).

Acuity scores serve as the target outcome for our precision modeling analysis. Using 1-year follow-up data from 105 youth tracked for adverse outcomes (incarceration, homelessness, residential placement, or psychiatric hospitalization), we developed a predictive model that categorizes youth into risk levels: 1 (no predicted adverse outcomes), 2 (1–2 predicted adverse outcomes), or 3 (3–4 predicted adverse outcomes). This model, which achieved 96.2% accuracy, was then applied to all study participants using their exit assessment data. The resulting predicted acuity scores became the outcome variable for identifying effective intervention combinations through precision modeling. Scores ≤ 1.6 indicate high likelihood of community stability for 12 months post-discharge.

The precision analytics model is trained using existing program administration data, which includes clinical assessments, trauma assessments, behavior assessments, thought pattern assessments, physicians' assessments, and incident data. Random Forest Regression with an ensemble of 100 individual decision trees is employed as the machine learning algorithm to identify matched comparison groups of children who are comparable across a range of factors. Our Random Forest models utilized 1,901 variables derived from Gemma Services' electronic medical record system, organized into three main categories: 1) context variables ($n = 164$) comprising clinical assessments at intake (Child and Adolescent Trauma Screen [CATS], Child PTSD Symptom

Scale [CPSS], Strengths and Difficulties Questionnaire [SDQ]); 15 diagnostic categories from the DSM-5; demographics including age, gender, and race that are used only for equity analysis, not matching; and system factors including payor source, expected discharge location, and length of stay; 2) experience variables (n = 1,603) calculated as average across time, change ratios, pre-post differences, slops, and total exposure, which encompass therapy modalities for individuals (40 techniques), families (42 techniques), groups (14 techniques), and DBT-specific interventions; 72 behavioral incident types tracked monthly; 32 shift-based milieu therapeutic interventions; and types and frequency of home/community visits; and 3) outcome variables (n = 122) including exit CATS, CPSS, and SDQ scores, discharge diagnoses, and predicted acuity scores. To capture treatment dynamics, we transformed each experience variable using eight different calculations (e.g., average monthly contacts, total duration, engagement levels, trauma depth addressed), resulting in the large variable count. We employed Random Forest regression with input shuffling to determine variable importance, as described by Shimshock et al. (2022). This approach systematically permutes each variable's values while holding others constant to measure impact on predictions, providing empirical justification for feature inclusion. Following variable importance ranking, we conducted feature-specific models to identify optimal "sweet spots" for the most important items, determining specific dosage levels and intervention thresholds associated with improved outcomes.

The matching process specifically excludes demographic characteristics or traits that should not influence treatment decisions, such as race, to minimize selection bias and ensure fair comparisons. The data includes records of different combinations of treatments, services, medications, events, and activities received by these matched groups. These groups are selected based on contextual and baseline intake characteristics that predict how likely a child/youth is to engage in programming promoting their ability to stay in the community post-discharge. While demographic characteristics are excluded from the matching process to avoid codifying differential treatment based on race or gender, these variables are rigorously analyzed in a subsequent stage. After identifying effective treatments for each matched group, we examine whether all demographic subgroups within each matched group equally received these interventions. This two-stage approach allows us to identify service delivery inequities without building assumptions about race-based treatment differences into our model. The use of effect coding in this analysis compares each demographic group to the overall average within their matched group, avoiding the problematic practice of using any racial group as a reference category.

After identifying treatments predictive of longer post-discharge tenures, inferential statistical analyses are conducted to compare outcomes for children who received the predicted services against those who did not. This process

relies on counterfactual analysis to determine the best course of action based on past successes. The model demonstrated strong predictive performance, with an overall accuracy of 86.8% in identifying which children would benefit from specific intervention combinations. Bootstrap sampling was used to create different training datasets for each tree in the Random Forest, leaving approximately one-third of the data "out-of-bag" for internal validation. Model performance was assessed using these out-of-bag samples. The F1 score, which balances precision and recall, was 85.8%, and the Area Under the ROC Curve (AUC) was estimated at 86.9%, reflecting strong discriminative ability well above the chance level of 50%.

These performance metrics suggest that the model may offer data-informed suggestions to support practitioner decision-making while maintaining appropriate boundaries for clinical judgment. Insight Generator was designed as a decision support tool rather than an automated decision-making system. Gemma Services practitioners maintain full autonomy to incorporate their clinical expertise and contextual factors that may not be captured in the model, using the tool's recommendations as one of the several inputs into their comprehensive assessment and treatment planning process.

A crucial step protecting against bias and the effects of the "coded gaze" involves assessing whether all children, regardless of their demographic background, received the interventions identified as most effective for their matched group. Rather than ignoring demographic factors entirely, the model evaluates potential bias after establishing predictive relationships by examining whether all demographic groups equally received what works. If outcome patterns vary across demographic groups, the Insight Generator can identify these differences by comparing outcomes within demographic blocks after matching on all relevant non-demographic variables. This two-stage approach (first determining what works through unbiased comparison groups, then analyzing for whom it works best) allows practitioners to identify and address service inequities while maintaining algorithmic transparency.

Additionally, the use of decision trees and random forest algorithms, which are less prone to overfitting compared to more complex "black-box" models, allows for better generalization, and the Insight Generator's human-in-the-loop approach further contributes to interpretability and provides opportunities for clinical oversight (see Figure 1 for workflow visualization). Input shuffling is used to systematically permute each variable's values while holding others constant and measuring the impact on predictions. This process provides empirical justification for features inclusion or exclusion and helps practitioners and families understand the relative contribution of each variable. The human-centered design approach involves practitioners, families, and children in the design and feedback process to help promote relevance and equity in the tool's insights.

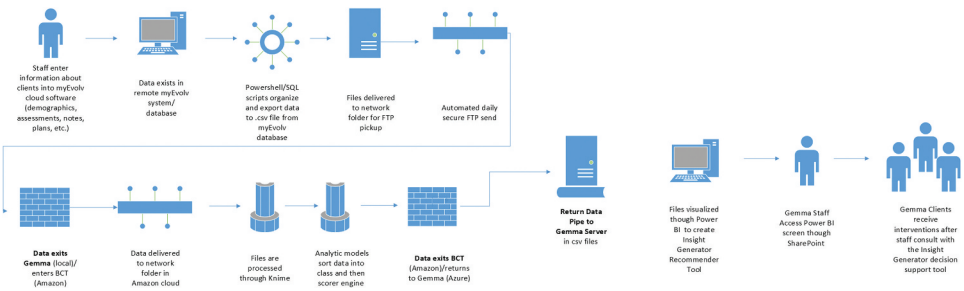

Figure 1. Workflow visualization.

Initial Insight Generator Implementation

In 2020, Gemma Services implemented the Insight Generator tool in the PRTF program to provide decision support to front-line staff, clinicians, and program leadership. This tool helps ensure continuity of care in high-turnover contexts like residential programs. All staff members receive a monthly e-mail with a report for their caseload that provides acuity scores predicting the likelihood of a child/youth staying in the community for 1 year if they were discharged that month, when the child/youth's most recent assessments were completed, and a summary of the child/youth's current needs and several strategy options that could be helpful in addressing these needs based on previous success. This delivery strategy allows any staff member to have access to the dashboards on any device on which they have an e-mail at any time of day.

The average intake acuity score is 2.17, with an average reduction in acuity of 29%, resulting in an average discharge acuity score of 1.87. These figures are based on a review of 163 children who exited the program, segmented into 8 six-month periods since January 2021. The smallest change in acuity was observed in the first six-month review, where initial acuity scores at intake were not recorded. Additionally, during the first 2 years of tool implementation (during the COVID-19 pandemic, when many providers were experience longer lengths of stay due to lockdowns), length of stay at Gemma Services was reduced by 29 days compared with the previous 2 years. A review of outcomes for the 4 years after initial tool implementation showed that children/youth served by Gemma have continued to have reductions in length of stay. It is important to note that a reduction in the number of referrals received from a primary referrer and statewide PRTF bed shortage occurred during the initial implementation period. Finally, the implementation has demonstrated feasibility across diverse populations. Initial equity analyses suggest variations in treatment outcomes by demographic groups within matched comparison groups (see Table 1). These preliminary findings require larger samples for definitive conclusions and are being monitored for quality improvement purposes.

Table 1. Model performance metrics.

Block Matched Group Propensity Score	Accuracy	F1 Score	Precision	Recall
0.140	0.884	0.737	0.778	0.700
0.290	0.774	0.741	0.833	0.667
0.400	0.867	0.824	0.778	0.875
0.491	0.895	0.889	0.889	0.889
0.515	0.879	0.889	0.941	0.842
0.818	0.909	0.945	0.963	0.929
Overall Model Performance	**0.872**	**0.882**	**0.843**	**0.862**

Evaluation of Gemma Services Practitioners Usage of Insight Generator

Researcher Characteristics and Positionality

The first author contracted with Gemma Services through a partnership with the second author, who helped lead design and implementation of the Insight Generator tool. Study funding covered the time and effort of the first author and the analysis team she leads (the third author is an undergraduate researcher, the fourth author is a postdoctoral scholar, and the fifth author is research staff). The second author does not supervise any Gemma Services staff members, and supervisors have no way of knowing which staff did and did not participate in the study.

Sampling Strategy

Our sampling strategy was based on the multiple perspective interviewing approach, in which viewpoints of members of the same social unit or organization are collected separately and triangulated in the analysis to explore interrelated roles and experiences with more privacy and confidentiality than joint interviews afford (Vogl et al., 2019). This approach also allows for the comparison of consonant and conflicting accounts across roles and perspectives (Vogl et al., 2019). Accordingly, focus groups were organized to convene practitioners with different roles in Gemma Services and Insight Generator implementation to gather insights across organizational positions while supporting psychological safety to honestly disclose experiences by placing staff and supervisors in separate groups. Groups were convened during regular team meeting times to reduce study burden. Initially, the team assumed that front-line unit staff were not using the tool, so groups were facilitated with only half of the front-line staff.

Data Collection Methods

The study employed Ripple Effects Mapping (REM), a group participatory evaluation method that engages program stakeholders with visual mapping to retrospectively review the chain of effects resulting from a complex program implementation with four components: 1) appreciative inquiry, 2) a participatory approach, 3) interactive group interviewing with reflection, and 4) mind mapping (Chazdon et al., 2017). For this project, after introducing the goal and obtaining consent, the first and second authors paired

participants for appreciative interviews in breakout rooms to share stories of how they have used information from the Insight Generator. Pairs reported on stories as the first author documented and thematically organized them using Xmind mind mapping software, and the second author "rippled out" the stories by asking probing questions such as: "What led to that happening?" "Then what happened?" and "Who was involved?" To enhance credibility, participants were asked to member-check mind-maps for accuracy. This study received an exemption certification (#93555) from the first author's institutional review board on 2/14/24.

Data Analysis Methods

The first author exported mind maps from Xmind into spreadsheets for analysis, and all five authors participated in an analysis meeting to topically group stories and classify topics as "facilitators" or "barriers" to effective Insight Generator implementation for presentation at two all-staff meetings. The combination of data from all groups preserved participant anonymity. At the all-staff meetings, the first and second authors presented topics and representative stories and asked participants to answer questions including how Gemma Services could better use the tool as a team; how Gemma Services might improve training and coaching on the tool; what is most interesting about the evaluation learnings; and what changes should be made in how the tool is used. These presentations facilitated additional member-checking of mind-maps and transcript quotations. The first and second authors also presented the topics and representative stories to Gemma Services leadership and gathered feedback.

Results

Table 2 displays the number of participants by role-based group. Participants were much more vocal about stories and experiences in initial groups than during the all-staff meeting, indicating the importance of facilitating role-based groups to mitigate the chilling effect that a supervisory presence may create. Figure 2 presents an example of the mind-maps generated by the initial group conversations, and Figure 3 summarizes the percentage of the overall dataset of stories that were coded to each topic.

Table 2. Participant type and number.

Participant Group	Number of Participants	Number of People in Role Invited to Participate	Proportion of Eligible Staff Participating
Unit Staff	8	46	17.4%
Therapists	7	8	87.5%
Supervisors	8	9	88.9%
Clinical Support Staff	5	5	100%
Leadership	3	3	100%
Total	31	71	43.7%

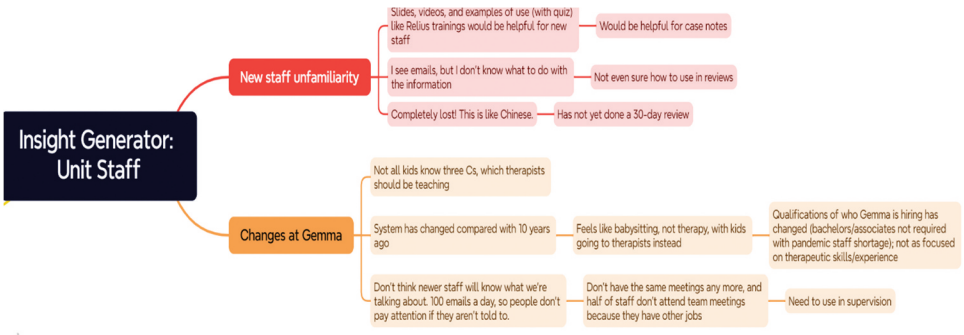

Figure 2. Example Mind Map.

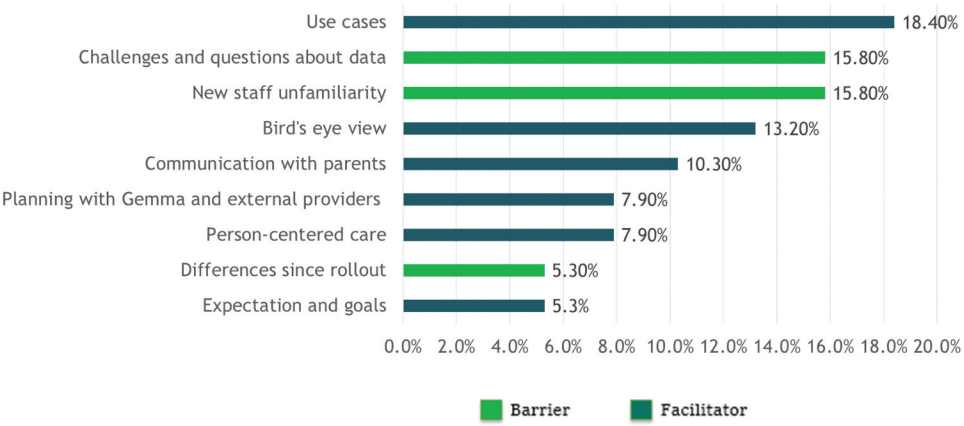

Figure 3. Frequency of stories by Topic.

Facilitators to Insight Generator Usage

The most common type of story participants shared about facilitators to tool implementation focused on use cases. For example, a supervisor talked about how one of their team members was applying Insight Generator to assess ongoing progress, saying, "So, he reads the information that comes out, and he's able to really look at the kid's progress over time and compare it to what's being seen on the unit. And he said, it's nice to be able to go back and look at what you have accomplished, right?" A therapist mentioned that the ability to track progress longitudinally also helps parents: "And a lot of times when parents are struggling, when they aren't seeing the progress, because they're not on campus all the time . . . being able to pull up the parenting version that I can share with them shows them the progress that has been made." The "parent version" of the Insight Generator provides visualizations tailored to caregivers showing what is going well and what needs ongoing work based on the combination of assessments and behavioral incidents to date. Supervisors reported using

Insight Generator in weekly supervision with therapists and in conversations with parents. The supervisor group also described how the tool has helped their teams take a "bird's eye view" of the discrepancies between how children/youth present on the unit and what is happening in therapy, which helps manage emotions around challenging discharge conversations because individual perceptions can be compared with the team's data. One supervisor shared that they used the tool to help their team navigate differences in opinions about a child/youth's treatment and that it led to better questions being asked about how a child/youth is doing.

Additionally, participants indicated the Insight Generator facilitates communication with parents and external partners, especially in preparation for discharge. For instance, a clinical support staff member said, "When [child] was ready for discharge and his dad was giving pushback, we were able to see when … was appropriate for him to discharge." Acuity scores help clinicians identify when the risk is lowest for youth to return home and are one of the factors considered when assessing readiness for a community setting. Another clinical support staff member cited the relevance of Insight Generator data and recommendations in coordinating services with external partners, such as schools and community-based providers. Leadership team members described using the tool to highlight the innovation and data-informed nature of the PRTF Program to external stakeholders, the joint commission during accreditation, and surveyors and auditors.

Contrary to expectations at the onset of the project that front-line unit staff did not use the tool, the evaluation surfaced stories of how unit staff drew upon Insight Generator e-mails to provide more person-centered care. One unit staff member shared that "there was a kid that when they had like a high behavior, they would go into the fight or flight response. So, [name] looked at the Insight Generator and tried a different way of approaching the child and actually using a different intervention to go about when that behavior occurs … It was just eye opening … It was like an 'aha' moment." Another staff member explained how they look at the monthly e-mail to see when a child/youth first came to Gemma Services, where they are now, and what areas need improvement and how they can help, saying "even if their acuity is either increasing or staying stagnant, the Insight Generator e-mail provides different things that we can use and try. And if that doesn't work, there's always another option." A supervisor described applying the tool to match areas of staff strength to particular child/youth based on tool recommendations, so that staff who are most skilled in delivering specific strategies are assigned to the child/youth for whom those strategies are suggested.

Barriers to Insight Generator Usage

Stories of barriers to Insight Generator utilization largely related to challenges and questions with the underlying data sources and new staff unfamiliarity.

For example, one of the therapists discussed their concern about the acuity score becoming a threshold or cutoff point, saying, "I just want to make sure that . . . it doesn't just become, 'Okay, this kid is under two, it's time for them to go,' right?" A supervisor echoed this concern, saying:

> I think one thing that's been a challenge for me was helping the therapist to view the results of the Insight Generator more as a progress report for them versus for the kids. Because the kids are not responsible for the interventions that are done for them. And people tend to focus on that acuity number and needing to get to a certain point and use it to try to motivate the kids. And that's not what it's for. It's to help the therapist and staff recognize that there are things that need to be done differently by us to get the kids to that point. But the kids are not going to just miraculously do it on their own because they were saying that they need to get there.

A few members of the therapist group noted that children/youth may not always answer the assessments accurately, as well as noted inconsistencies they have observed between acuity scores and behavior. Staff mentioned that not all behaviors are documented and therefore included in Insight Generator, particularly during mornings that are often chaotic.

Another barrier to effective Insight Generator utilization after initial rollout with intensive training in 2020 is assuring that new staff members are trained on appropriate applications of the tool. One of the unit staff groups explained that new staff may not be told about or trained on the Insight Generator, that they receive a high volume of other e-mails, that the reports are confusing, and that many of the staff have a second job and are unable to attend the team meetings. A unit staff member said they felt like people were speaking a foreign language when they heard about the Insight Generator. As one unit staff member described it, "we have a hundred e-mails a day. So, if nobody says you need to read this, nobody's even going to know about it . . . So, I think when they have a meeting with the supervisor, they should be going over this stuff." Another unit staff member said that integration of the tool into their work requires practice: "Like we've been practicing on how to use it, so that when people ask me questions, I feel more comfortable without having to call [second author] to answer the staff's question. So just us being in it and having access to it. And we have access to it, but like really using it as a tool that comes up when we are problem-solving interventions." Members of different groups reflected on changes to broader organizational practices and hiring patterns since initial Insight Generator rollout, reporting that cross-team meetings that previously had leveraged the tool are no longer regularly occurring and that workforce strains in the wake of the pandemic have led to less rigorous hiring qualifications that make tool uptake more challenging for new hires.

Strategies to Improve Tool Usage

Sharing these facilitators and barriers during the all-staff and leadership report-out meetings led to identification of strategies to improve usage of the Insight Generator. First, there was consensus that the tool should be integrated into daily practices across roles, especially at the supervisory level, as opposed to having staff use it at their discretion. Based on this feedback, leadership reported they "worked a lot with the Residential Life Managers this summer to push them into being a little bit more comfortable . . . One of the things that someone suggested that I really liked was the little pieces, like not making the Insight Generator take over a whole meeting or be the whole topic, but have it be a two-minute, 'Oh, let's see what they say. Okay, now what are we looking at?' to really kind of just have it as . . . a tool in our toolbox, not like our whole meeting is now 45 min of looking at this data." Another quality improvement recommendation that emerged from discussion was to recognize and celebrate front-line staff use of the tool for person-centered care and time logged into the system. Finally, participants discussed enhancing training strategies, like creating modules in their online training system with slides, videos, and examples of use by roles with quizzes to support rollout of the next version of Insight Generator as well as encouraging informal peer mentoring. A member of the leadership team suggested that "we reach out to the staff . . . that we know do know how to use it and be like how do we get more regular time for you to use it? And then we're asking them, okay, can you pick one other staff in your unit and go teach them how you're using this?" One participant mentioned that the REM evaluation approach, particularly the appreciative interview, helped facilitate sharing of knowledge and experiences with the tool.

Discussion

The Insight Generator contributes to improved outcomes for children in residential treatment by recommending tailored interventions specifically for the children and families being served, moving away from a "one-size-fits-all" approach to ensure that each child receives the interventions most likely to be effective for their specific circumstances. The Insight Generator empowers practitioners identify and administer person-centered interventions, avoiding the risks of over- or under-treatment. Additionally, by assessing each child's trauma within the context of their life and tailoring care accordingly, the tool aims to facilitate a compassionate, respectful treatment process. The Insight Generator's design focused on equity assures the benefits of the decision support tool are distributed fairly across all children in care. By optimizing therapeutic outcomes and advancing understanding of effective, equitable care pathways, the tool contributes to a more informed, data-driven approach to

treatment planning and delivery, with initial findings showing that tool usage at Gemma Services allowed children to return to their communities an average of 29 days sooner.

The Insight Generator is designed to be iteratively remodeled based on additional data. Findings from this study and other feedback from the first 4 years of implementation are currently being leveraged by Gemma Services to improve Version 2 of the tool. First, the updated version adding more closed cases will incorporate recommendations for service arrays, dosage, and ideal length of stay for each comparison group. Additionally, Version 2 will be more inclusive of neurodiversity and applicable across programs as well as incorporate available contextual data on social drivers of health and family resources. Several dashboard improvements are in development, including unit-specific and caseload views. With the learnings from Version 2, Gemma Services will identify areas where more data are needed, and data from all closed cases will help inform future versions.

Other residential providers can apply these types of AI-based tools by developing internal capacity and/or contacting outside organizations to help complete this work. Successful implementation requires a close working relationship between clinicians, data analysts, and programmers to ensure the insights generated are validated and meaningful in practice. The implementation must include ongoing training and use support to maintain implementation of the tool as designed. The second author and BCT Partners are available to brainstorm with other providers who are interested in exploring adoption of the tool.

This study is limited by the potential of selection bias introduced by the relatively low participation rate of unit staff compared with staff in other roles and social desirability bias for staff to provide answers and responses that support the utility of the Insight Generator to other Gemma Services team members. Additionally, we hope to include family and youth experiences with Insight Generator co-design and implementation in future studies.

Yet, this study has transferable implications for the broader use of AI-based decision support tools in residential interventions. Continual user experience assessment across roles is important when implementing AI-based decision support tools in residential interventions to ensure staff understand why and how data are being sourced and to integrate tool use cases into routines of practice. Additionally, turnover and changes in hiring qualifications as well as organizational practices necessitate continual training and communication about decision support tools among supervisors for consistency. This continual communication should convey that decision support assessments represent a progress report for practitioners, not children/youth and are never a substitute for their judgment and decision-making. Finally, despite concerns

that AI-based decision support tools may perpetuate systemic discrimination, usage of the Insight Generator with its guardrails against black-box algorithms demonstrates the potential of these tools to help practitioners resist biased, patterned responses to behaviors and children/youth exhibiting them with varied person-centered recommendations. Using tools like the Insight Generator can help remind practitioners that, as one unit staff member said, "there's always another option."

Acknowledgments

We would like to thank the participants and leadership of Gemma Services for their time and insights. We would also like to thank BCT Partners, Peter York (pyork@bctpartners.com) and Stephen Shimshock (sshimsock@bctpartners.com) for their time and support in helping to share the use of the Insight Generator.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Gemma Services and the University of Kentucky College of Public Helath Undergraduate Research Scholarship.

References

Barnett, E. R., Boucher, E. A., Daviss, W. B., & Elwyn, G. (2018). Supporting shared decision-making for Children's complex behavioral problems: Development and user testing of an option GridTM decision Aid. *Community Mental Health Journal*, *54*(1), 7–16. https://doi.org/10.1007/s10597-017-0136-5

Bauer, G. R., & Lizotte, D. J. (2021). Artificial intelligence, intersectionality, and the future of public health. *American Journal of Public Health*, *111*(1), 98–100. https://doi.org/10.2105/AJPH.2020.306006

Buolamwini, J. (2023). *Unmasking AI: My mission to protect what is human in a world of machines*. Random House.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science (new York, NY)*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Chatpreecha, P., & Usanavasin, S. (2023). Design of a collaborative knowledge framework for personalised attention deficit hyperactivity disorder (ADHD) treatments. *Children*, *10*(8), 1288. https://doi.org/10.3390/children10081288

Chazdon, S., Emery, M., Hansen, D., Higgins, L., & Sero, R. (2017). *A field Guide to ripple effects mapping*. University of Minnesota Libraries Publishing. http://conservancy.umn.edu/handle/11299/190639

Chor, K. H. B., Epstein, R. A., & Luo, Z. (2023). Developing and validating a predictive risk Model for youth placement in residential care to support decision-making under the family

first prevention services act. *Residential Treatment for Children & Youth*, *40*(3), 324–347. https://doi.org/10.1080/0886571X.2022.2111018

Clausen, C., Leventhal, B., Nytrø, Ø., Koposov, R., Røst, T. B., Westbye, O. S., Koochakpour, K., Frodl, T., Stien, L., & Skokauskas, N. (2023). Usability of the IDDEAS prototype in child and adolescent mental health services: A qualitative study for clinical decision support system development. *Frontiers in Psychiatry*, *14*, 1033724. https://doi.org/10.3389/fpsyt.2023.1033724

Dettlaff, A. J., & Boyd, R. (2022). The causes and consequences of racial disproportionality and disparities. In R. D. Krugman & J. E. Korbin (Eds.), *Handbook of child maltreatment* (pp. 221–237). Springer International Publishing. https://doi.org/10.1007/978-3-030-82479-2_10

Elkington, K. S., Wasserman, G. A., Ryan, M. E., Sichel, C. E., Sarapas, C., Dennis, M. L., & Taxman, F. S. (2023). E-Connect: Linking probation youth at risk for suicide to behavioral health services. *Journal of Consulting and Clinical Psychology*, *91*(9), 547–557. https://doi.org/10.1037/ccp0000824

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Gay, K. E. M., & York, P. (2018). *A new way to use data: Precision care for better outcomes in psychiatric residential treatment for children*. https://www.scattergoodfoundation.org/wp-content/uploads/2018/02/A-New-Way-to-Use-Data.pdf

Hall, S. F., Sage, M., Scott, C. F., & Joseph, K. (2023). A systematic review of sophisticated predictive and prescriptive analytics in child welfare: Accuracy, equity, and bias. *Child & Adolescent Social Work Journal*, *41*(6), 831–847. https://doi.org/10.1007/s10560-023-00931-2

LaBrenz, C. A., Littleton, T., Shipe, S. L., Bai, R., & Stargel, L. (2023). State policies on child maltreatment and racial disproportionality. *Children and Youth Services Review*, *151*, 107048. https://doi.org/10.1016/j.childyouth.2023.107048

McIntosh, A., Lyons, J. S., Weiner, D. A., & Jordan, N. (2010). Development of a Model for predicting running away from residential treatment among children and adolescents. *Residential Treatment for Children & Youth*, *27*(4), 264–276. https://doi.org/10.1080/0886571X.2010.520634

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Shimshock, S., Chor, K. H. B., & Brylske, P. D. (2022). Using latent class analysis to identify clinical subgroups and pathways of youth in a therapeutic foster care program. *Children and Youth Services Review*, *141*, 106626. https://doi.org/10.1016/j.childyouth.2022.106626

Stewart, S. L., Celebre, A., Hirdes, J. P., & Poss, J. W. (2020). Risk of suicide and self-harm in kids: The development of an algorithm to identify high-risk individuals within the Children's mental health system. *Child Psychiatry & Human Development*, *51*(6), 913–924. https://doi.org/10.1007/s10578-020-00968-9

Theall, L., Arbeau, K., Willoughby, K., St. Pierre, J., Ng, G., & Stewart, S. L. (2022). Assessing needs and outcomes of children and youth receiving intensive services. *Residential Treatment for Children & Youth*, *39*(1), 16–33. https://doi.org/10.1080/0886571X.2021.1964415

Troy, J. D., Torrie, R. M., & Warner, D. N. (2021). A machine learning approach for identifying predictors of success in a Medicaid-funded, community-based behavioral health program using the child and adolescent needs and strengths (CANS). *Children and Youth Services Review*, *126*, 106010. https://doi.org/10.1016/j.childyouth.2021.106010

Trudeau, K. J., Yang, J., Di, J., Lu, Y., & Kraus, D. R. (2023). Predicting successful placements for youth in child welfare with machine learning. *Children and Youth Services Review*, *153*, 107117. https://doi.org/10.1016/j.childyouth.2023.107117

Van Der Put, C. E., Bouwmeester-Landweer, M. B. R., Landsmeer-Beker, E. A., Wit, J. M., Dekker, F. W., Kousemaker, N. P. J., & Baartman, H. E. M. (2017). Screening for potential child maltreatment in parents of a newborn baby: The predictive validity of an instrument for early identification of parents at risk for child abuse and neglect (IPARAN). *Child Abuse & Neglect*, *70*, 160–168. https://doi.org/10.1016/j.chiabu.2017.05.016

Vogl, S., Schmidt, E.-M., & Zartler, U. (2019). Triangulating perspectives: Ontology and epistemology in the analysis of qualitative multiple perspective interviews. *International Journal of Social Research Methodology*, *22*(6), 611–624. https://doi.org/10.1080/13645579.2019.1630901

York, P. (2021). *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*. Project Evident. https://projectevident.org/case-study/generating-on-demand-actionable-evidence-first-place-for-youth-and-gemma-services%ef%bf%bc/