

An Artificial Intelligence-Driven System to Predict ASD Outcomes in ABA

Dr. David J. Cox, Ph.D., MSB, BCBA-D^{1,2}, Dana D'Ambrosio, MS, BCBA, LBA¹,
Jamie Pagliaro¹, Rethink Data Team¹

¹ Rethink First; ² Endicott College

Past researchers have sought to describe and predict how individuals with autism spectrum disorders (ASD) are likely to benefit from applied behavior analysis (ABA) therapy. These studies, however, have had limited generalizability due to sample sizes, simple modeling approaches, and failing to include more holistic patient profiles. Further, few studies have embedded their results into technology platforms practitioners can incorporate into treatment settings. In this article, we provide an overview of how we used 48 variables spanning hours and characteristics of ABA, treatment goals characteristics, and patient characteristics to predict goals mastered for 31,294 individuals with ASD receiving from 615 companies. Unsupervised machine learning identified between 8-56 distinct patient clusters (depending on the algorithm) differing along characteristics known to influence outcomes from past published research and patient characteristics unpublished in past research predicting patient progress from ABA. Linear regression models (as used in past research) led to an overall r^2 value ($r^2=.90$; MAE=1.30) that was $\sim.23$ higher than previously published studies. Machine learning improved predictions further ($r^2=.99$; MAE=1.04). When predictions were made within patient clusters, r^2 ranged between .95-.99 ($\sim.20$ -.24 points higher than past research) and MAE ranged between 1.12-1.45. To close, we describe how this Artificial Intelligence (AI) system is embedded within a technology platform continuously collecting data. This allows the system to improve over time and, in turn, users of the AI system can use the results to improve ASD outcomes through use cases such as: (a) real-time recommendations of ABA dosage based on unique patient characteristics; (b) feedback on actual versus expected patient outcomes; and (c) how patient progress varies along social determinants of health. In the future, these data and the underlying models could be leveraged by payors and providers alike to support and enable their unique value-based care initiatives.

Keywords: outcomes; artificial intelligence; machine learning; autism; applied behavior analysis

Predictability is a common concern in the delivery of healthcare services. Patients typically want to know how likely it is that each of several treatment options will be effective (e.g., Meyer et al., 2021); and they want to know the potential treatment length and cost relative to their personal budget and calendar so they can plan and budget accordingly (e.g., Bhargava & Loewenstein, 2015; Han et al., 2012). Likewise, insurance companies often want to know how potential treatment length and cost so they can forecast expenses and plan strategically for the future (e.g., Papanicolas et al., 2018). Finally, healthcare practitioners want to know potential treatment length, the possible range

in which outcomes can improve, and the resources it will require so they can manage things such as caseloads, waitlist communication, hiring, and business operations (e.g., WHO, 2015).

It is through this lens that the current work sits relative to one area of healthcare in the delivery of Applied Behavior Analysis (ABA) services for individuals with autism spectrum disorders (ASD). The prevalence of ASD diagnosis has increased in the previous 20 years from 1 in 150 to 1 in 44 children (Centers for Disease Control and Prevention, 2021). As a result, stakeholders (patients, parents, payors, and providers) are increasingly aware of the variability in cost, treatment duration, and patient outcomes. For example, the average annual cost of treatment is estimated to range anywhere between \$17,227 to \$130,182 (Autism Speaks, 2022; Eldevik et al., 2009); and individuals diagnosed with autism and their families might receive ABA for between 5 to 40 hours per week for a duration ranging between 18 months to five years (Larsson, 2012). Such variability in cost and duration of treatment following

Conflict of Interest: The AI system (patent pending) and products leveraging this AI system are the property of RethinkFirst.

Funding: No external funding was used to support the work published herein.

Disclaimer: The AI system described herein is continuously ingesting more data and the modeling methods are being continuously refined to improve performance. Thus, all results discussed herein should be considered as the static state of the system at the time of writing in February, 2023.

an ASD diagnosis adds significant uncertainty to all involved (De Groot & Thurik, 2015; Han et al., 2012; Meyer et al., 2021).

Compounding this uncertainty around increasing amounts of ABA service delivery is a lack of standardization for ABA dosage as a function of patient presentation at intake and throughout the duration of services. At the time of writing¹, the process of determining optimal dosage of ABA for each patient is seemingly subjective. Past research suggests a patient's age, symptom severity, historical treatment duration, personal/family needs, current abilities, and the overall goals of therapy *should* play a role in how many hours per week of ABA an individual might need (e.g., APBA, 2020; Granpeesheh et al., 2009). However, no known quantitative methodology with robust empirical support allows clinicians to translate those variables into a precise recommendation of the optimal number hours of ABA. Thus, a tool that provides reliable and precise predictions based on patient data and published best-practice evidence would help decrease the subjectivity and uncertainty around ABA treatment.

In addition to decreased subjectivity and reduced uncertainty around ABA treatment, board certified behavior analysts (BCBAs) arguably have an ethical obligation to improve in this area. The Behavior Analyst Certification Board (BACB) *Ethics Code for Behavior Analysts* includes several guidelines that speak directly to this topic (BACB, 2020). For example, Guideline 2.08 requires discussing the scope of treatment with a client before starting services; Guideline 3.01 requires BCBAs to identify and act upon opportunities that lead to avoidable harm or wasteful allocation of resources; and, Guideline 3.12 requires BCBAs to advocate for appropriate services. Here, leveraging a reliable, precise, and evidence-based method to recommend ABA dosage allows the BCBA to accurately communicate the scope of likely treatment; why they believe benefits will be maximized while avoiding wasteful allocation of resources; and to know that the amount and level of behavioral services they are recommending is data-based.

Past researchers have sought to improve predictability and aid BCBAs in their ethical obligations by quantitatively modeling the relationship between the hours that patients contact

ABA and the resulting outcomes. For example, Linstead and colleagues (2017a) modeled the relationship between treatment intensity and mastered learning objectives for 726 children aged 1.5-12 years who received community-based behavioral intervention services. Linear regression and neural network models led to r^2 values of .35 and .60, respectively. Similarly, in a separate study by Linstead et al. (2017b), they used linear regression to model the relationship between treatment intensity/duration and goals mastered within specific skill domains for 1,468 individuals with ASD. Across goal domains, r^2 values ranged between .50 (social goals) and .67 (motor goals). As a final example, Ostrovsky and colleagues (2022) used a series of t -tests, effect size measures, and Pearson correlations to quantify the relationship between changes in standardized assessment scores as a function of hours of ABA and modality of supervision. For the 178 individuals included in the study, they observed that clinically significant improvements in function were independent of the hours of ABA received.

Though broader limitations to this (and other) past studies are described in more detail below, the variance accounted for by these models was likely impacted by several factors. First, the overall sample sizes in these studies were relatively small². Second, as noted by the authors of several studies, models have often been built without accounting for the heterogeneity that comprises the broad spectrum that is ASD and that was unlikely to be fully captured in the studied samples. A final limitation to past work in this realm (noted by Ostrovsky et al., 2022) is the assumption of a linear relationship between hours per week of ABA and progress made. Dose-response curves in other areas of behavioral healthcare are often decidedly nonlinear (e.g., Dews, 1955; Levy et al., 2020; Zoladz & Diamond, 2009). Thus, modeling approaches that more flexibly account for nonlinear relationships may perform better (e.g., the neural network outperforming linear regression in Linstead et al., 2017a).

PREDICTING PATIENT PROGRESS BASED ON PATIENT & ABA CHARACTERISTICS

One technique that researchers have used to create data-driven subgroups from larger datasets

¹ February, 2023.

² We want to reiterate that this is a *relative* claim as they are among some of the largest studies to date examining these kinds of

relationships. "Small" here is in reference to the total number and amount of heterogeneity present with the ASD population as a whole and compared to the sample size in the current study.

is called cluster analysis (e.g., Ansari et al., 2018; Charron et al., 2023; Hewlett et al., 2022). Cluster analysis at its simplest can be defined as finding groups in data (Hennig et al., 2016). For our purposes here, cluster analyses would seek to separate “all individuals with ASD in our sample” into smaller subgroups based on the totality of their demographics and clinical-developmental states while they contact ABA. The idea is that modeling relationships between contact with ABA and patient progress might either be more accurate when conducted within patient clusters or lead researchers to differentially identify important variables that predict patient progress for specific clusters. In contrast to traditional modeling techniques where the relations between independent variables and a dependent variable is known, cluster analyses are often more exploratory as the researcher may not know how many clusters should exist nor which patients should be grouped in which cluster. Uncovering these relationships is often the goal of the analysis.

Researchers have used cluster analysis techniques to identify subgroups of individuals within samples of data gathered from individuals with ASD. For example, Parlett-Pelleriti and colleagues (2022) published a review of how unsupervised machine learning has been used to identify patient clusters for individuals with ASD. Table 1 shows the descriptive statistics of the sample sizes and number of clusters identified across the 36 studies included. Two items are of note, here. First, considering the heterogeneity of ASD, the sample sizes were relatively small with the median sample size being 220, 78% of the studies used less than 1000 participants, and all but one used less than 5000 participants. It is unknown what the clustering results might look like when a larger sample of individuals with ASD are included.

The second item to note were the variables used for clustering analyses and the resulting number of clusters. The median number of clusters identified across studies was three with results largely clustering around the ASD diagnostic criteria or assessment domains. This makes sense as those were the data used for clustering in many of the studies. But individuals with ASD are much more than the developmental and behavioral patterns contained in diagnostic or skills-based assessments. Minimally, they may have comorbid diagnoses or medical concerns that influence their unique clinical course (e.g., Lingren et al., 2016). Additionally, individuals with ASD (like the rest of us) are embedded within larger socioeconomic,

Table 1. Descriptive statistics of research applying unsupervised machine learning to cluster individuals with ASD (Parlett-Pelleriti, 2022).

	Sample Size	Number of Clusters
Maximum	20,658	7
Arithmetic Mean	1,218	3.32
Standard Deviation	3,504	1.21
Median	220	3.00
Minimum	10	2.00
Current Work*	31,294	8-53

* NB: Because this is a “live” system, the overall sample size is continuously increasing and the specific clustering results are continuously being refined toward optimal.

familial, and educational contexts. Past researchers have found that many variables outside of diagnostic symptomatology will predict response-to-intervention such as health-related features of neighborhoods and socioeconomic factors (e.g., Braverman & Gottlieb, 2014). Clustering analyses that attempt to identify more holistic subgroups of ASD beyond diagnostic and assessment classifications will need to include a greater number of variables.

A second limitation to past research unmentioned thus far has been how a stakeholder might use the information from the published analyses. For example, practitioners might be interested in answers to questions such as: How might the models provide a recommended dosage of ABA for an individual? How might practitioners design treatment programs based on this information? How could practitioners gain feedback around the patients’ outcomes they observe and what is expected based on these models? And, when considering the entirety of a patients’ unique human situation, what variables are most predictive of intervention success, and which are least predictive? Lastly, how might answers to these questions be packaged into a system that makes access to clustering and predictive models as efficient as possible?

The purpose of this manuscript is to describe at a high-level how researchers have been addressing the abovementioned limitations while attempting to provide answers to the practical practitioner questions. To do this, we first provide an overview of a patent-pending Artificial Intelligence (AI) based system that: continuously collects data on patient and ABA-related variables with a known relation to patient progress from ABA therapy; aggregates the data for analysis; uses unsupervised machine

learning to cluster patients and supervised machine learning to predict goals mastered; and returns the model outputs back into a technology platform to be used by an end-user. Following this high-level overview, we describe the general results of patient clustering analyses and how predicting patient outcomes is impacted when using these patient characteristics. In total, this manuscript provides a description of how the AI system functions, continuously improves, and preliminary results which will likely continue to improve.

HIGH-LEVEL SCHEMATIC OF THE AI SYSTEM

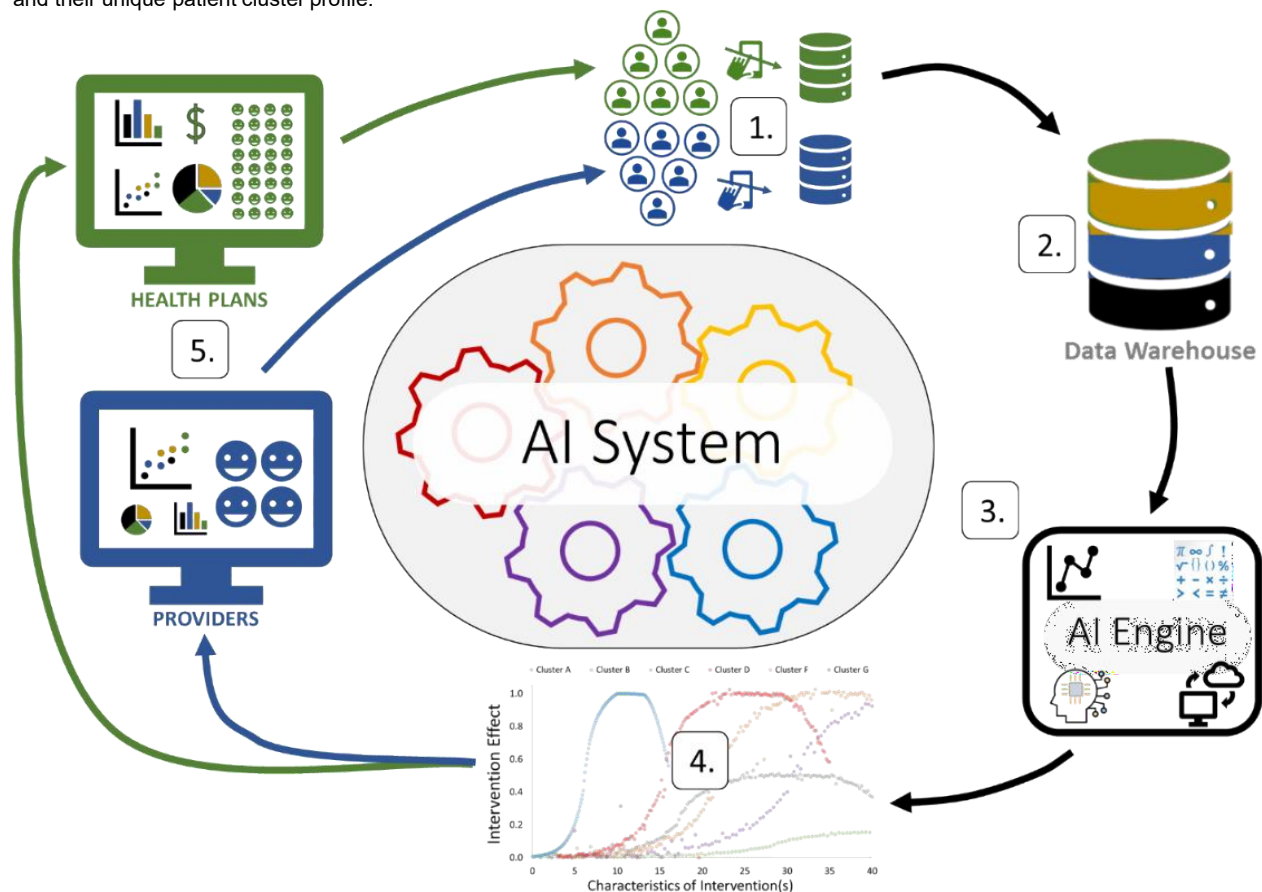
Figure 1 shows a high-level overview of the patent-pending AI system. The system involves five major steps. First, users interact with a web-based platform to enter the raw data for the patient and ABA-related variables used in the AI Engine (see below for more details). The specific application that captures these variables can differ depending on the variable, end-user, and product they are using. The output of this first step is raw data having

been collected and stored in a product-specific database.

The second step of the AI system is to use extract, transform, and load (ETL) workflows to move the data from the product-specific databases associated with its collection to a single, central location (i.e., a data warehouse). Specifically, the ETL workflows act on the output of the first step and regularly move data from the product specific databases to a centralized data warehouse within the scope of a data model allowing for the data from many of the tables to be related to one another. The primary output of the second step is a set of relational tables where each table contains some portion of the total set of variables used for modeling in the AI engine.

The third step involves a series of Python scripts that conduct the data pre-processing and analytics referred to as the AI Engine. The AI Engine is described in more detail below. Here, the primary points are that the Python scripts involved in this workflow: (a) pre-process the data stored in the relational tables in the data warehouse; (b)

Figure 1. High-level schematic of the AI system to predict patient clusters and patient outcomes based on contact with ABA therapy and their unique patient cluster profile.



combine the variables from many related tables into a single analytic data frame; (c) conduct unsupervised machine learning relative to patient clustering; (d) conduct a suite of statistical and supervised machine learning analyses relative to predicting patient outcomes using the patient-related and ABA-related variables; and (e) deploy two top-performing models as API endpoints—one each for patient clustering and predicting outcomes. The primary output of this third step are the mathematical-computational models for patient clustering and predicting patient outcomes.

The fourth step uses the primary outputs of step three to build dose-response curves unique to individual patients. Specifically, a user interacts with a web-based tool to provide data specific to the patient-related and ABA-related variables needed for the finalized models from step 3. Once collected, that data can be run through the clustering and outcome prediction models to generate a unique dose-response curve predicting estimated progress as a function of the patient characteristics and a range of hours of ABA the patient could contact.

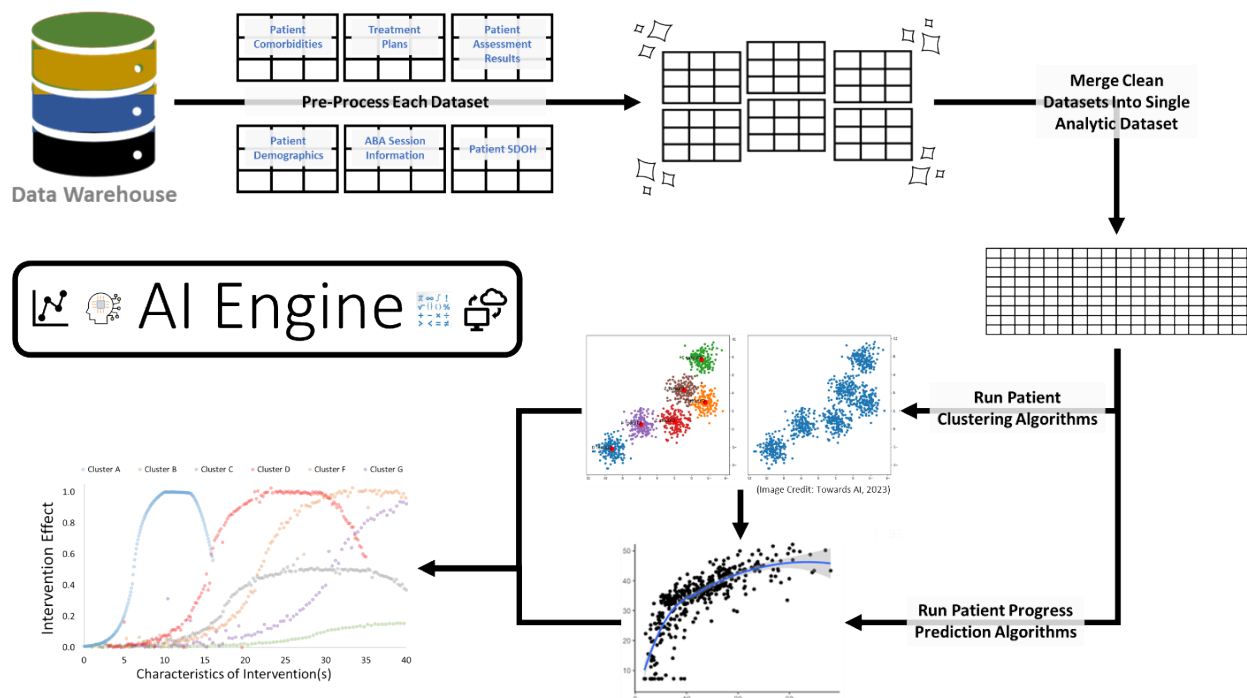
The fifth step uses the dose-response created in step four toward some practical aim. Few people are likely interested in identifying patient clusters or predicting outcomes simply for the sake of doing so. Rather, they want to use the information to help them do something better than they currently do. To

highlight what’s possible, we walk through three possible use cases based on the output of step four in the final section below. These include: (a) using the dose-response curve to recommend hours per week of ABA to optimize patient progress; (b) to gain performance feedback between observed and expected patient outcomes for a clinician’s caseload or an organization as a whole; and (c) how patient progress varies along social determinants of health when accounting for the total milieu of patient characteristics and their interactions. Before we get to these use cases, however, we will first briefly review how the AI Engine works and current results around patient clustering and precision predicting patient progress.

HIGH-LEVEL OVERVIEW OF THE AI ENGINE

As noted above, the AI Engine is comprised of five distinct steps (Figure 2). These are: (a) pre-process the data stored in the relational tables in the data warehouse, (b) integrate the various tables into a single, analytic data frame, (c) conduct unsupervised machine learning relative to patient clustering; (d) conduct a suite of statistical and supervised machine learning analyses relative predicting patient outcomes using the patient-related and ABA-related variables; and (e) deploy two top-performing models as API endpoints—one

Figure 2. High-level overview of the steps that comprise the AI Engine portion of the AI system.



each for patient clustering and predicting outcomes. In this section we provide a high-level overview of what occurs in each section. Of note, manuscripts that provide significant detail for steps (c) and (d) are in preparation and will be published in the near future.

Variable Selection. We sought to create a comprehensive set of variables that capture patient characteristics spanning diagnostic characteristics (DSM-5), patient age and history of treatment, clinical presentation of social and behavioral excesses and deficits, medical and behavioral health comorbidity, goal counts related to developmental domain subsets, social drivers of health, and barriers to progress/treatment. The 48 variables chosen for inclusion in the current analyses were derived from published standards (e.g., Autism Commission on Quality, 2022; Behavioral Health Center of Excellence, 2022; National Autism Center, 2009) and applicable research literature that outline best practices. Once selected, a team of experienced BCBA's and BCBA's at the doctoral level (BCBA-Ds) narrowed and rounded out the list of variables for inclusion which were then solidified by a provider advisory committee to ensure agreement on the likely relation between included variables and patient outcomes.

Pre-processing. Pre-processing is a series of steps that convert the data from its raw form stored in the data warehouse to a dataset on which statistical and machine learning analyses can be conducted (for reviews on methods, see Baskar et

al., 2013; Müller & Guido, 2017; Saleem et al., 2014). Generally, pre-processing involves data cleaning, data integration, data transformation, and dimensionality reduction. These steps and the strategies used within the current AI system are shown in Table 2. Data cleaning often, at minimum, involves employing explicit strategies around handling missing data, outliers, and inconsistent data. Data integration involves joining or merging datasets together after aggregating each dataset at the level required for analyses. Here, data were aggregated and integrated at the individual patient level spanning one week of ABA service delivery. Data transformation refers to a set of mathematical techniques whereby data are converted from their original values to values likely to improve subsequent modeling (Müller & Guido, 2017). For example, data transformations might include: generalization (e.g., converting from zip code to state), normalization (e.g., log-transforms, min-max scaling), feature reduction (e.g., removing variables with little variability; combining features that are similar in some way); or feature discretization (i.e., converting continuous data into bins or categories).

Unsupervised ML: Patient Clustering. Unsupervised machine learning refers to a suite of mathematical and algorithmic techniques aimed at learning the underlying structure in data where the answer is unknown (Patel, 2019). Clustering refers to a suite of mathematical and algorithmic techniques to find groups in data (Hennig & Meila, 2016). Combined, the goal of unsupervised machine learning for patient clustering is to identify patient subgroups within the larger dataset that are

Table 2. Overview of the sequence and strategies used in pre-processing data to prepare it for patient clustering analyses and to predict patient outcomes.

Step	Strategy
1.	Drop patients and features where the amount of missing data would likely impact their reliability and are likely to also impact model precision (e.g., patients with fewer than 100 sessions, features with 60% or more missing data).
2.	Identify and Winsorize outliers.
3.	Impute missing data by chained equation (Buck, 1960; van Buuren & Groothuis-Oudshoorn, 2011).
4.	Use zip code for feature engineering (e.g., add neighborhood and economic SDOH proxies; convert to US region and one-hot encode)
5.	Aggregate and quantify patient developmental and familial characteristics (e.g., age, number of siblings).
6.	Aggregate intervention characteristics across skill acquisition domains.
7.	Aggregate intervention characteristics across reduction targets, intensities, and functional assessments.
8.	Aggregate and quantify medical and behavioral health comorbidities.
9.	Aggregate and quantify patient progress, generally, as well as across specific skill domains.
10.	Aggregate and quantify hours per week of ABA contact per patient for the duration they are in the dataset.
11.	Merge all datasets together aggregated at the level of patient-week (e.g., continuous variables convert to arithmetic mean; categorical convert to one-hot encoded; ordinal convert to median).
12.	Examine how aggregation influenced feature distributions and modify where significantly impacted.
13.	Normalize feature scales using min-max to 0-1.

similar when considering the 48 variables that comprise the final analytic dataset.

Table 3 shows the patient clustering ($N=31,294$) fit metrics for four algorithms wherein each takes a unique definition for how to define a cluster³. Overall, maximum Silhouette Coefficients ranged between 0.25-0.47, maximum Calinski-Harabasz indices ranged between 6,086.66-13,718.52, and optimal patient clusters ranged between 8-to-56 depending on the algorithm and level of granularity a researcher or practitioner is interested in obtaining or is practically useful for them. Figure 3 shows a three-dimensional plot of the clustering results for agglomerative hierarchical clustering wherein the underlying assumptions are likely best justified based on the function of clustering patients with ASD in the current context (for a comprehensive treatment of these assumptions see Hennig et al., 2016). Each color in the plot represents a unique patient cluster and each marker represents a unique patient.

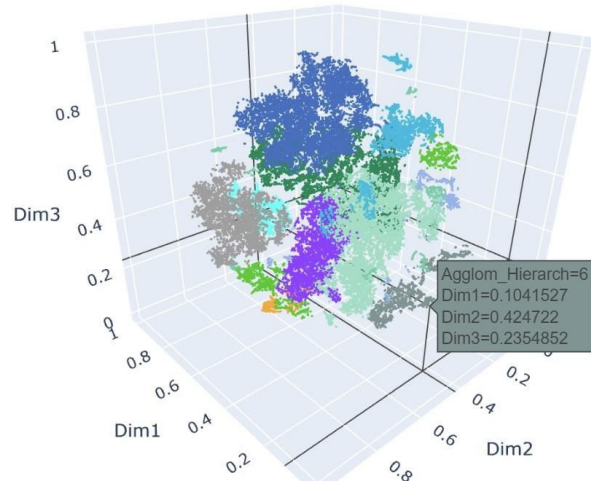


Figure 3. Example patient cluster visual using agglomerative hierarchical clustering. Each color represents a unique cluster as identified by the algorithm. Each individual marker represents a single patient.

A natural question is exactly how each of these patient clusters differ from each other. Figure 4 shows 10 of the individual patient characteristics included in the current analysis. Five of those patient characteristics are variables that have been associated with differential patient outcomes following contact with ABA in past research (top five panels). Five of the patient characteristics in Figure 4 have no known previous publications associating the variable with differential patient outcomes following contact with ABA at the time of this writing (bottom five panels).

There are two important observations to be made based on the results above and considering the scope of this current manuscript. First, the clustering algorithms consider all 48 variables simultaneously with each cluster being a unique combination of all patient characteristics. So, the ten variables shown in Figure 4 should be considered a demonstration of what this system currently allows us to do. The current results should be considered preliminary and will continue to improve as the AI system collects more data and our methodologies are further refined. Second, a follow-up manuscript is in preparation that provides significantly greater detail around the patient

Table 3. Overview of the best clustering fit metrics for the algorithms tested ($N=31,294$). For both metrics, a higher score indicates better clustering.

Clustering Algorithm	Silhouette	Calinski-Harabasz	Optimal Clusters
k-Means	0.25	8,169.60	30
Agglomerative Hierarchical	0.47	13,718.52	8
HDBSCAN	0.37	2,377.78	56
BIRCH	0.39	6,086.66	18

Silhouette equals the ratio of average dissimilarity of each observation with its own cluster (a) relative to the average dissimilarity to observations in the closest other cluster (b)(Halkidi et al., 2016). This can be written as an equation: $s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}$. *Calinski-Harabasz equals the ratio of the sum of between-cluster dispersion and of within-cluster dispersion* (Halkidi et al., 2016). As an equation: $CH(C_k) = \frac{\text{trace}(\sum_{j=1}^K n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T)}{\text{trace}(\sum_{j=1}^K \sum_{i \in (j)} (\bar{x}_i - \bar{x}_j)(\bar{x}_i - \bar{x}_j)^T)} \times \frac{n-K}{K-1}$. *In more lay terms, Silhouette coefficients provide a sense of how well an observation belongs to its assigned cluster compared to any other cluster; and Calinski-Harabasz indices indicate how well partitioned the clusters are.*

³ The reader is referred to Hennig et al. (2016) for a review of the assumptions and underlying

mathematics used to define a cluster via each algorithm.

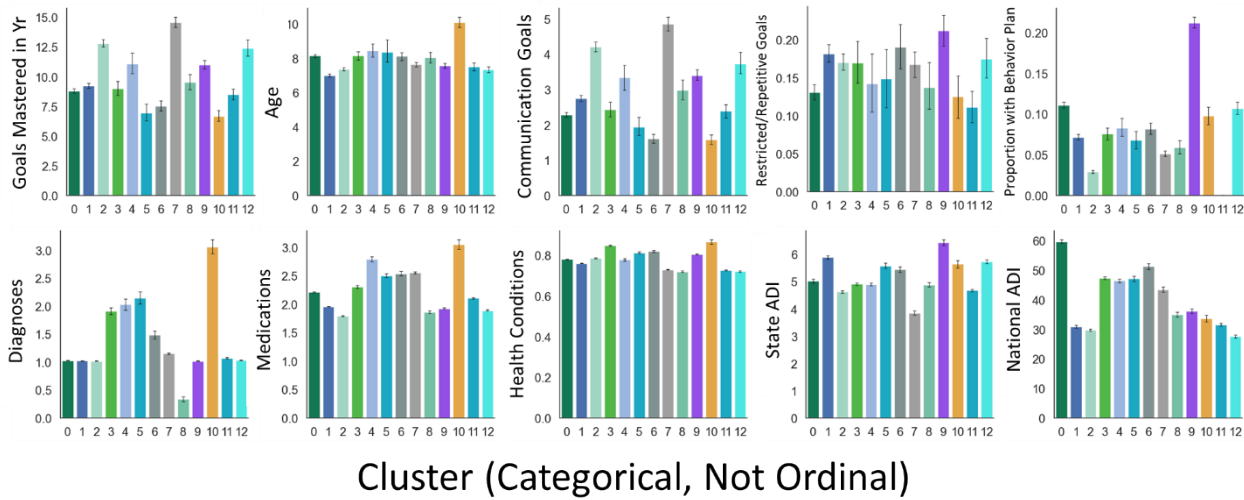


Figure 4. Demonstration of patient characteristics that differ among agglomerative hierarchical clusters. The top panel shows differences in variables previously published in the literature. The bottom panels show differences in variables related to health that could logically impact progress.

clustering methods, their evaluation, and how these results were derived. Again, the scope of this manuscript is to highlight the totality of the system and what it can do. The first author can be contacted for specific methodological questions in the interim.

Supervised Machine Learning: Predicting Patient Outcomes. Supervised machine learning refers to a suite of mathematical and algorithmic techniques aimed at learning the underlying structure in data where the answer is known (Müller & Guido, 2017). For our purposes, the answer we know and want to predict is how many goals each patient mastered during a given period of time. The goal of supervised machine learning is to create a mathematical equation (or set of equations) that relates the patients-related and ABA-related characteristics to how many goals they mastered. Once built, the model can be used to predict how many goals a new patient is likely to master based on their unique set of characteristics.

Different statistical and machine learning algorithms range from less (linear regression) to more (ensemble models) complex. Less complex models are often more parsimonious though can sometimes be less accurate than more complex models. In contrast, more complex models are often more accurate though, due to their complexity, can sometimes be more difficult for users to intuitively understand how the input variables relate to the predicted output. As a result of the tradeoff between model simplicity

and complexity, we analyzed a series of increasingly complex statistical and machine learning models for how well they could predict patient progress as a function of contact with ABA and each patient’s unique characteristics.

Table 4 shows an overview of the fit metrics when assessing each algorithm using 10-fold cross-validation with 80%-20% test-train splits. That is, we started by first randomly splitting our data into two datasets. One dataset was used to train the model (80% of all observations) and a holdout dataset was used to test the predictive capabilities of the model (20% of all observations). We then repeated this process nine more times using a different selection of test-train splits of the dataset with each iteration. Table 4 shows the arithmetic mean of the fit (r^2) and loss (mean absolute error; MAE) metrics for each of the 10-folds for each of the algorithms using their default parameters.

Overall, the random forest regressor algorithm led to the lowest MAE values across

Table 4. Average fit and loss metrics from 10-fold cross-validation using the test data sets not included in model training.

Algorithm	r^2	MAE
Linear Regression	0.90	1.30
k -Nearest Neighbors	0.91	1.27
Support Vector Regressor	0.99	1.06
Random Forest	0.99	1.04
AdaBoost	0.88	1.32

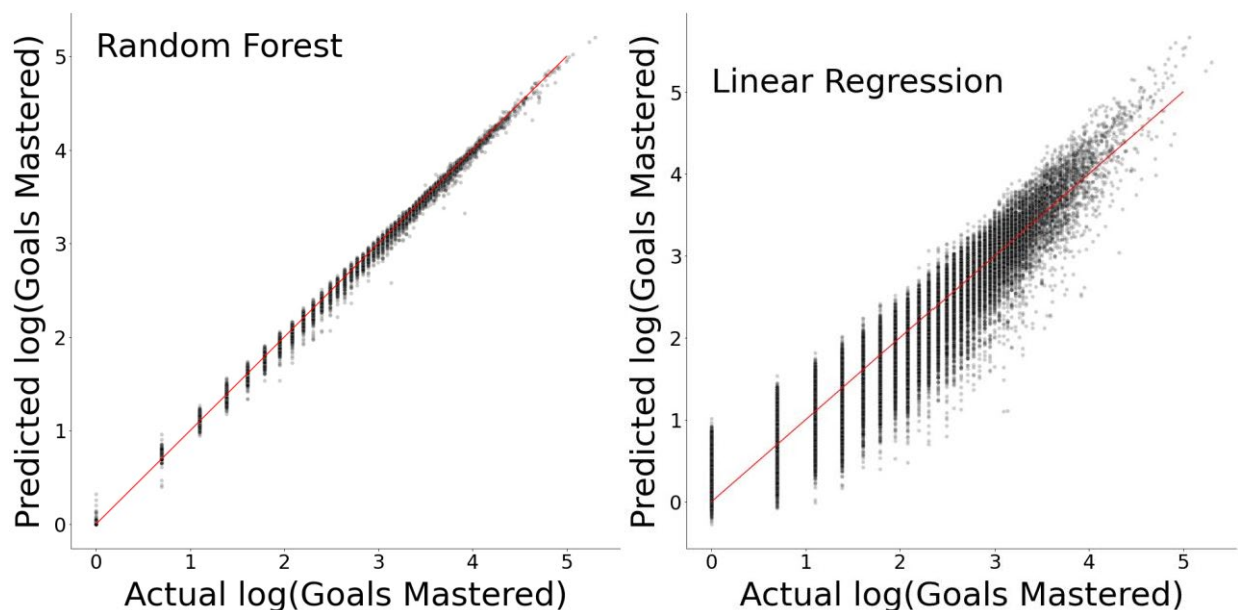
both training and test datasets. To build a final predictive model, we conducted hyperparameter tuning via grid search across model parameters of the number of trees and the maximum depth allowed. The left panel in Figure 5 shows the results of using the final model built using the random forest algorithm. The final fit and loss metrics of the top performing model with the holdout test data were $r^2 = .99$ and MAE = 1.12. The right panel in Figure 5 shows the results of using linear regression similar to researchers in past published studies. Fit and loss metrics using linear regression were $r^2 = .90$ and MAE = 1.82, respectively. Of note, the obtained r^2 using linear regression is .23 points higher than the highest known previous model predicting goals mastered by individuals with ASD as a function of contacting ABA (Linstead et al., 2017b reported an r^2 of .67 for motor goals, MAE was unreported). The predictive model using machine learning was .32 points higher.

We conducted the same set of analyses for each patient cluster in the dataset. That is, we first isolated the data to only the patients in a given cluster, assessed each algorithm using 10-fold cross-validation with 80%-20% test-train splits, conducted hyperparameter tuning via grid search to identify a top model, and then made predictions of patient progress on the holdout test set. When predicting goals mastered within unique clusters, r^2 values ranged from .95-.99 and MAE values ranged 1.12-1.45. Of note, the r^2

values obtained here are ~.20-.24 higher than the cluster specific r^2 values observed by past researchers which ranged from .64 to .75 (Stevens et al., 2019). Thus, the AI system approach taken in this work appears to be a significantly more precise method for predicting patient progress compared to previously published research. This appears true, overall, as well as within unique patient clusters.

As with the section on patient clustering, there are two important observations to be made based on the results above and considering the scope of this manuscript. First, this system continuously receives more patient data and we are actively identifying better ways to collect and incorporate data spanning a greater number of variables. Thus, the results in this section should be considered a demonstration of what this system currently allows us to do. The current results should be considered preliminary and will continue to improve as the AI system collects more data and our methodologies are further refined. Second, a follow-up manuscript is in preparation that provides significantly greater detail around using machine learning to predict patient outcomes across goals mastered and changes in VBMAPP scores. Again, the scope of this manuscript is to highlight the totality of the system and what it can do. The first author can be contacted for specific methodological questions in the interim.

Figure 5. Predicted vs. observed goals mastered for the models derived using the random forest algorithm (left panel) and using linear regression (right panel). In each panel, the red line shows where exact matches between observed and predicted occur.



Deploy as API endpoint. The last phase in the AI Engine is to make the final, top-performing models available to users of a technological platform. Deploying machine learning models can be accomplished in a number of ways (see Treveil et al., 2020 for an introductory text). For the technology ecosystem at RethinkFirst, Microsoft Azure makes it easy to deploy machine learning models as a REST API endpoint. Users with access to this technology system can then call the API endpoint with the requisite data fields and have the model return a value. What value gets returned, however, would vary with the use case and how the API endpoint is embedded within the product.

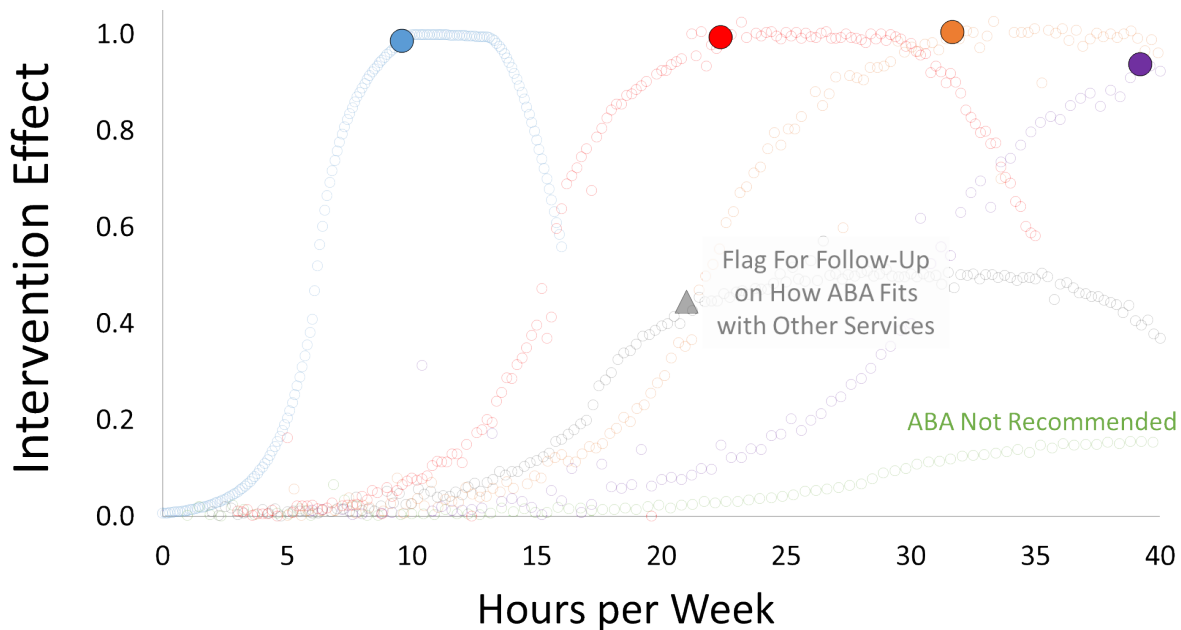
USE CASES

Recommending ABA Dosage. A straightforward application of the finalized model would be to have the model recommend the optimal hours per week relative to goals mastered. Figure 6 shows hypothetical data wherein patient characteristics can be passed to the predictive model along with a range of hours of ABA the patient might contact. Such dose-response curves are commonly used to identify optimal therapeutic effect of an intervention in behavioral pharmacology (e.g., Dews, 1955) and have been used previously for analyses specific

to ABA (e.g., Ostrovsky et al., 2022; Stevens et al., 2019). For each patient, there is likely to be a point wherein little-to-no further progress would be predicted if more hours of ABA were contacted (the large filled-in circles in Figure 6). This point of little-to-no further return could then be the recommended hours per week of ABA.

Observed vs. Expected Patient Outcomes. Another straightforward application of the finalized model would be to retroactively compare a patient’s actual progress to expected progress. Specifically, a patient’s unique characteristics could be passed to the model along with the hours of ABA they contacted. The model would return the number of goals expected to be mastered based on other patients who present with similar characteristics and received similar amounts of ABA. By comparing the observed patient progress to the expected patient progress, the provider or their supervisor can get feedback on how well they are providing services with current clients. In turn, providers can follow-up with specific patients and the supervising BCBAs to identify why they are achieving less than expected outcomes. And, for patients for whom the provider is observing greater than expected outcomes, providers can follow-up to identify if there are variables predicting better success that they can use with other patients to further improve outcomes. Further, these data and the underlying models could be leveraged by payors

Figure 6. Example use case wherein the predictive model produces an expected dose-response curve based on hours per week of ABA along with the patient characteristics that determine cluster assignment. Each color would be a unique dose-response curve predicted for a patient based on their unique profile.



and providers alike to support and enable their unique value-based care initiatives.

Associated Social Determinants of Health.

Many aspects of our social environment have been repeatedly shown to play a role in patient outcomes and are referred to as social determinants of health (e.g., Braverman & Gottlieb, 2014). Because several of these variables are included in the model, the user could analyze these data to help identify and potentially mitigate the effects of these variables on patient outcomes. For example, similar to the previous use case, a provider could systematically identify how well each patient in their organization or on their caseload is performing relative to expected outcomes. The provider could then determine whether differences between observed and expected outcomes were associated with unique variables known to be a social determinant of health (SDOH; e.g., race, neighborhood walkability, income, education). By identifying specific SDOH that impact the patients on their caseload or in their organization, providers would better know how to identify resources for those clients that might mitigate this effect and improve their overall outcomes. Similar to feedback on provider performance, analyses related to SDOH could be leveraged by payors and providers alike to support and enable their unique value-based care initiatives.

LIMITATIONS AND FUTURE DIRECTIONS

There are several limitations to the current patent-pending AI system and underlying AI Engine, each of which points to future directions for building upon and improving the AI system as described herein. The first limitation is that the patients included in this work do not comprehensively account for all possible combinations of unique individual characteristics that an individual with ASD might present to ABA with (i.e., sampling bias). Nevertheless, the current study did contain data from 31,294 individuals with ASD making it the largest known published study on individuals with ASD at the time of writing⁴. Thus, though all research and technological systems will necessarily be incomplete in terms of the included sample (Jennings & Cox, 2023), the sample size in the present work is 1.51 times larger than the next

closest study (Lingren et al., 2016), 6.35 times larger than the third largest study (Doshi-Velez et al., 2014), and 142.25 times larger than the median sample size used in previous cluster analyses for individuals with ASD (Parlett-Pelleriti, 2022). For past work predicting patient outcomes as a function of ASD, the current study is 21.32 times larger than the sample size of the second largest study (Linstead et al., 2017b).

A second limitation and area for future direction relates to the variables chosen for inclusion in patient clustering and for predicting patient progress. We used domain expertise and past published research to identify and include variables with a logical or known relationship to patient progress following contact with ABA. However, behavior analysts have historically underreported demographic variables within the clinical literature (e.g., Jones et al., 2020) and have only recently begun to consider how SDOH might influence outcomes (e.g., Wright, 2022). Further, it is possible that a different approach to aggregating ABA session characteristics and treatment plans would lead to better models—work we are feverishly conducting. As noted in the introduction, the purpose of this manuscript was to showcase how this system currently operates and its current effectiveness. Future work using this AI system will include further exploration of variable inclusion and feature engineering.

A final area for future research involves more direct measures around how the outputs of this AI system impact clinical decision-making and stakeholders more broadly. As shown in Figure 1, the output of this system can inform a variety of products for a variety of end users. By examining how end users behave before and after use of the AI system, future research will be able to identify if and how clinician decision-making changes as a function of the system and how those decisions might influence patient progress. Outside of the AI system and end products themselves, however, there are likely to be additional, unmeasured variables that interact with this information to multiply influence the decisions behavior analysts make. Without knowledge of what those variables might be, collaboration with end users will be crucial for identifying if and where those additional variables might need to be added to the AI system as a whole.

⁴ February, 2023.

SUMMARY

All stakeholders alike would benefit from the ability to predict likely patient progress as a function of contact with ABA therapy and the patient's unique characteristics. Past researchers have sought to describe and predict how individuals with ASD are likely to benefit from ABA therapy. However, the results of these studies have had limited generalizability due to sample sizes, simple modeling approaches, and a lack of including robust patient cluster analyses to individualize patient predictions. Further, few studies have directly converted the insights gained into a technology that others could incorporate into their treatment settings to improve patient outcomes. Above, we described how we have built and are leveraging a patent-pending AI system that: integrates the continuous and ongoing collection of patient data in treatment outcomes and patient characteristics; allows us to identify robust patient clusters unique from previous literature given its significantly larger sample size; combines patient cluster characteristics with ABA hours per week to predict within 1.12 goals mastered within a 12-month ABA treatment period; how the models can be deployed back into a technological system for user interaction; and three subsequent use cases in: (a) real-time recommendations of ABA dosage based on unique patient characteristics; (b) feedback on actual patient outcomes relative to expected outcomes; and (c) how patient progress varies along social determinants of health.

REFERENCES

Ansari, W. E., Ssewanyana, D., & Stock, C. (2018). Behavioral health risk profiles of undergraduate university students in England, Wales, and Northern Ireland: A cluster analysis. *Frontiers in Public Health*, 6, 120. <https://doi.org/10.3389/fpubh.2018.00120>

Association of Professional Behavior Analysts (2020). ABA Autism Guidelines: Applied Behavior Analysis Practice Guidelines for Autism Spectrum Disorder. Retrieved from: <https://casproviders.org/wp-content/uploads/2020/03/ABA-ASD-Practice-Guidelines.pdf>

Autism Commission on Quality (2022). ACQ Applied Behavior Analysis Accreditation program standards and guidelines (1.0). Retrieved from: <https://autismcommission.org/standards/>

Autism Speaks (2022). Autism statistics and facts. Retrieved from: <https://www.autismspeaks.org/autism-statistics-asd>

Baskar, S. S., Arockiam, L., & Charles, S. (2013). A systematic approach on data pre-processing in data mining. *Compusoft: An International Journal of Advanced Computer Technology*, 2(11), 335-339.

Behavior Analyst Certification Board (2020). *Ethics code for behavior analysts*. <https://bacb.com/wp-content/ethics-code-for-behavior-analysts/>

Behavioral Health Center of Excellence (2022). 2022 standards and new accreditation model. Retrieved from: <https://www.bhcoe.org/2022-standards-and-new-accreditation-model/>

Bhargava, S., & Loewenstein, G. (2015). Choosing a health insurance plan: Complexity and consequences. *Journal of the American Medical Association*, 314(23), 2505–2506. <https://doi.org/10.1001/jama.2015.15176>

Braverman, P., & Gottlieb, L. (2014). The social determinants of health: It's time to consider the causes of the causes. *Public Health Reports*, 129(Suppl 2), 19-31. <https://doi.org/10.1177%2F00333549141291S206>

Centers for Disease Control and Prevention (2021). Data & Statistics on Autism Spectrum Disorder. Retrieved from: <https://www.cdc.gov/ncbddd/autism/data.html>

Charron, E., Yu, Z., Lundahl, B., Silipigni, J., Okifuji, A., Gordon, A. J., ... & Cochran, G. (2023). Cluster analysis to identify patient profiles and substance use patterns among pregnant persons with opioid use disorder. *Addictive Behaviors Reports*, 17, 100484. <https://doi.org/10.1016/j.abrep.2023.100484>

De Groot, K., & Thurik, R. (2018). Disentangling risk and uncertainty: When risk-taking measures are not about risk. *Frontiers in Psychology*, 9, 2194. <https://doi.org/10.3389/fpsyg.2018.02194>

Dews P. B. (1955). Studies on behavior. I. Differential sensitivity to pentobarbital of pecking performance in pigeons depending on the schedule of reward. *The Journal of Pharmacology and Experimental Therapeutics*, 113(4), 393–401.

Doshi-Velez, F., Ge, Y., & Kohane, I. (2014). Co-morbidity clusters in autism spectrum disorders: an electronic health record timeseries analysis. *Pediatrics*, 133(1), e54–e63.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Using participant data to extend the evidence base for intensive behavioral intervention for children with autism. *Journal of Autism and Developmental Disorders*, 39(11), 1302–1310. <https://doi.org/10.1007/s10803-009-0732-7>

Granpeesheh, D., Dixon, D. R., Tarbox, J., Kaplan, A. M., & Wilke, A. E. (2009). The effects of age and treatment intensity on behavioral intervention outcomes for children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 3(4), 1014-1022. <https://doi.org/10.1016/j.rasd.2009.06.007>

Halkidi, M., Vazirgiannis, M., & Hennig, C. (2016). Method-independent indices for cluster validation and estimating the number of clusters. In: C. Henning, M. Meila, F. Murtagh, & R. Rocci (Eds.), *Handbook of Cluster Analysis* (pp. 595-618). Taylor & Francis Group. ISBN: 978-1-4665-5189-3

Han, P. K. J., Klein, W. M. P., & Arora, N. K. (2012). Varieties of uncertainty in health care: A conceptual taxonomy. *Medical Decision Making*, 31(6), 828-838. <https://doi.org/10.1177%2F0272989X11393976>

Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (2016). *Handbook of Cluster Analysis*. Taylor & Francis Group. ISBN: 978-1-4665-5189-3

- Hennig, C., & Meila, M. (2016). Cluster analysis: An overview. In: C. Hennig, M. Meila, F. Murtagh, & R. Rocci (Eds.), *Handbook of Cluster Analysis* (pp. 1-20). Taylor & Francis Group. ISBN: 978-1-4665-5189-3
- Hewlett, M. M., Raven, M. C., Graham-Squire, D., Evans, J. L., Cawley, C., Kushel, M., & Kanzaria, H. K. (2022). Cluster analysis of the highest users of medical, behavioral health, and social services in San Francisco. *Journal of General Internal Medicine*. <https://doi.org/10.1007/s11606-022-07873-y>
- Jennings, A., & Cox, D. J. (2023). Starting the Conversation Around the Ethical Use of Artificial Intelligence in Applied Behavior Analysis. *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/zmn28>
- Jones, S. H., St. Peter, C. C., & Ruckle, M. M. (2020). Reporting of demographic variables in the *Journal of Applied Behavior Analysis*. *Journal of Applied Behavior Analysis*, 53(3), 1304-1315. <https://doi.org/10.1002/jaba.722>
- Larsson, E. V. (2012). Applied behavior analysis (ABA) for autism: What is the effective age range for treatment? *The Lovaas Institute for Early Intervention Midwest Headquarters White Paper*. Retrieved from: <https://childisplayautism.com/wp-content/uploads/Effective-Treatment-Range-1.pdf>
- Levy, H. C., Worden, B. L., Davies, C. D., Stevens, K., Katz, B. W., Mammo, L., Diefenbach, G. J., & Tolin, D. F. (2020). The dose-response curve in cognitive-behavioral therapy for anxiety disorders. *Cognitive Behaviour Therapy*, 49(6), 439-454. <https://doi.org/10.1080/16506073.2020.1771413>
- Lingren, T., Chen, P., Bochenek, J., Doshi-Velez, Manning-Courtney, P., Bickel, J., ..., & Savova, G. (2016). Electronic health record based algorithm to identify patients with autism spectrum disorder. *PLoS ONE*, 11(7), e0159621. <https://doi.org/10.1371/journal.pone.0159621>
- Linstead, E., Dixon, D. R., French, R., Granpeesheh, D., Adams, H., German, R., ..., & Kornack, J. (2017a). Intensity and learning outcomes in the treatment of children with autism spectrum disorder. *Behavior Modification*, 41(2), 229-252. <https://doi.org/10.1177/0145445516667059>
- Linstead, E., Dixon, D. R., Hong, E., Burns, C. O., French, R., Nocak, M. N., & Granpeesheh, D. (2017b). An evaluation of the effects of intensity and duration on outcomes across treatment domains for children with autism spectrum disorder. *Translational Psychiatry*, 7(9), e1234. <https://doi.org/10.1038%2Ftp.2017.207>
- Meyer, A. N. D., Giardina, T. D., Khawaja, L., & Singh, H. (2021). Patient and clinician experiences of uncertainty in the diagnostic process: Current understanding and future directions. *Patient Education and Counseling*, 104(11), 2606-2615. <https://doi.org/10.1016/j.pec.2021.07.028>
- Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python*. O'Reilly Media, Inc. ISBN: 978-1-449-36941-5
- National Autism Center. (2009). National Standards Project: Phase 1. Retrieved from: <https://nationalautismcenter.org/national-standards/phase-1-2009/>
- National Autism Center (2015). National Standards Project, Phase 2. Randolph, MA: National Autism Center. Retrieved from: <https://www.nationalautismcenter.org/national-standards/>
- Ostrovsky, A., Willa, M., Cho, T., Strandberg, M., Howard, S., & Davvitian, C. (2022). Data-driven, client-centric applied behavior analysis treatment-dose optimization improves functional outcomes. *World Journal of Pediatrics*. <https://doi.org/10.1007/s12519-022-00643-0>
- Papanicolaou, I., Woskie, L. R., & Jha, A. K. (2018). Health care spending in the United States and other high-income countries. *Journal of the American Medical Association*, 319 (10), 1024-1039. <https://doi.org/10.1001/jama.2018.1150>
- Patel, A. A. (2019). *Hands-on unsupervised learning using Python: How to build applied machine learning solutions from unlabeled data*. O'Reilly Media, Inc. ISBN: 978-1-492-03564-0
- Parlett-Pelleriti, C. M., Stevens, E., Dixon, D., & Linstead, E. J. (2022). Applications of unsupervised machine learning in autism spectrum disorder research: A Review. *Review Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s40489-021-00299-y>
- Saleem, A., Asif, K. H., Ali, A., Awan, S. M., & Alghamdi, M. A. (2014). Pre-processing methods of data mining. *IEEE/ACM 7th International Conference on Utility and Cloud Computing*, 451-456. <https://doi.org/10.1109/UCC.2014.57>
- Stevens, E., Dixon, D. R., Novack, M. N., Granpeesheh, D., Smith, T., & Linstead, E. (2019). Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervised machine learning. *International Journal of Medical Informatics*, 129, 29-36. <https://doi.org/10.1016/j.ijmedinf.2019.05.006>
- Towards AI (2023). Centroid neural network : An efficient and stable clustering algorithm. Retrieved from: <https://towardsai.net/p/centroid-neural-network-an-efficient-and-stable-clustering-algorithm>
- Treveil, M., & Dataiku Team (2020). *Introducing MLOps: How to Scale Machine Learning in Enterprise*. O'Reilly Media, Inc. ISBN: 978-1-492-08329-0
- World Health Organization (2015). Health systems financing: The path to universal coverage. Retrieved from: https://apps.who.int/iris/bitstream/handle/10665/155002/WHO_HIS_SDS_2015.6_eng.pdf
- Wright, P. (2022). Social determinants of health and applied behavior analysis: Leveraging the connection to make the world a better place. CEU Event available from: <https://behavioruniversity.com/social-determinants-health-aba>
- Zoladz, P. R., & Diamond, D. M. (2008). Linear and non-linear dose-response functions reveal a hormetic relationship between stress and learning. *Dose-Response: a publication of International Hormesis Society*, 7(2), 132-148. <https://doi.org/10.2203/dose-response.08-015.Zol>