Published in final edited form as:

J Appl Behav Anal. 2018 July ; 51(3): 443–465. doi:10.1002/jaba.477.

IDENTIFYING PREDICTIVE BEHAVIORAL MARKERS: A DEMONSTRATION USING AUTOMATICALLY REINFORCED SELF-INJURIOUS BEHAVIOR

Louis P. Hagopian,

THE KENNEDY KRIEGER INSTITUTE AND JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE

Griffin W. Rooker,

THE KENNEDY KRIEGER INSTITUTE AND JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE

Gayane Yenokyan

JOHNS HOPKINS UNIVERSITY BLOOMBERG SCHOOL OF PUBLIC HEALTH

Abstract

Predictive biomarkers (PBiOMs) are objective biological measures that predict response to medical treatments for diseases. The current study translates methods used in the field of precision medicine to identify PBiOMs to identify parallel *predictive behavioral markers (PBMs)*, defined as objective behavioral measures that predict response to treatment. We demonstrate the utility of this approach by examining the accuracy of two PBMs for automatically reinforced self-injurious behavior (ASIB). Results of the analysis indicated both functioned as *good to excellent* PBMs. We discuss the compatibility of this approach with applied behavior analysis, describe methods to identify additional PBMs, and posit that variables related to the mechanisms of problem behavior and putative mechanism of treatment action hold the most promise as potential PBMs. We discuss how this technology could guide individualized treatment selection, inform our understanding of problem behavior and mechanisms of treatment action, and help determine the conditional effectiveness of clinical procedures.

Keywords

automatically reinforced self-injurious behavior; conditional effectiveness; conditional probability analysis; precision medicine; predictive behavioral markers

The use of functional analysis (FA) to inform treatment selection for problem behavior is well established and represents current best practices. Once the function of problem behavior is identified there are a myriad of function-based treatment options available, however, there is little research to indicate which treatment from among those options might

be the most effective for a given case. For example, functional communication training and noncontingent reinforcement are both well-established treatments (Carr, Severtson & Lepper, 2009; Kurtz, Boelter, Jarmolowicz, Chin, & Hagopian, 2011), but we do not know circumstances under which one might be more effective than the other. Identifying variables (in addition to the function of behavior) that predict response to treatment could further enhance behavior analysts' ability to select treatments that are the most likely to be effective for a given case. We can look to the paradigm of precision medicine and its use of predictive biomarkers to provide a model for advancing a more individualized approach to treatment selection.

Just as behavior analytic approaches attempt to look beyond the manifest features of the presenting problem (i.e., response topography) and seek to understand it at the level of its causal mechanisms (i.e., response function), so does the paradigm of precision medicine. Precision medicine has emerged with advances in genomics and other technologies that enable researchers to identify disease mechanisms and other factors that impact both disease susceptibility and treatment outcomes (National Research Council, 2011). In addition to guiding treatments that are highly individualized, precision medicine has shifted disease classification from a taxonomy based on signs and symptoms to one based on causal mechanisms (Katsnelson, 2013). The overarching purposes of this article are to demonstrate how some methods and concepts used within precision medicine are translatable to applied behavior analysis, and to suggest that this approach has the potential to advance behavioral-analytic research and practice. We first provide a brief overview of some of the concepts and methods used in precision medicine, and we then apply these methods to identify predictive behavioral markers (PBMs) for automatically reinforced SIB (ASIB).

PRECISION MEDICINE: CONCEPTS AND METHODS

Precision Medicine

Precision medicine relies in part on identification of disease-relevant biomarkers, which are defined generally as *objective biological measures that are related to, a product of, or the cause of a disease* (La Thangue & Kerr, 2011). Biomarkers include genetic mutations that cause the disease or affect how medications targeting diseases are metabolized, epigenetic changes that affect gene expression that may trigger diseases, and levels of a substance in the blood or tissue related to the disease. Biomarkers can be used in different ways. For example, *diagnostic biomarkers* identify a disease, including subtypes of a disease; *prognostic biomarkers* provide information about the future course of a disease and how it is likely to advance; and *predictive biomarkers* predict response to a specific treatment (Drucker & Krapfenbauer, 2013). Some biomarkers can be used in multiple ways. For example, the presence of mutations in BRCA1 gene is both a prognostic biomarker for a subtype of ovarian cancer, and a predictive biomarker for response to chemotherapy (Carser et al., 2011).

Predictive biomarkers (PBioMs) represent a specific type of biomarker most relevant to the current study. A PBioM's utility is based on how accurately it can distinguish responders from nonresponders to a particular treatment. Once a PBioM's accuracy is demonstrated, and its generality is validated through replication, it is then used in clinical practice to guide

the selection of treatments that are best matched to the individual patient. The number of identified PBioMs continues to grow, perhaps most rapidly in oncology in which several kinds of cancers have been identified as having subtypes, some of which have been found to respond differentially to specific types of treatments (La Thangue & Kerr, 2011). For example, non-small-cell lung cancer is a type of lung cancer with known subtypes, one of which is characterized by mutations in the EGFR gene. Medications targeting that specific mutation (EGFR inhibitors) have been shown to be both (a) more effective for treating that cancer subtype relative to other subtypes and (b) more effective than standard chemotherapy for that subtype (Okimoto & Bivona, 2014). As a result of these findings, patients with non-small-cell lung cancer now routinely undergo genetic testing to identify the type of mutation that is present so that they can receive the most effective treatment sooner (and forgo less effective treatments that might also be more invasive and have adverse side effects).

Some parallels between precision medicine and behavior-analytic assessment and treatment of problem behavior should be immediately evident. The use of PBioMs to guide the selection of treatments that are matched to causal variables underlying the presenting problem mirrors how behavior-analytic interventions are designed based on the function of the behavior rather than its topographical properties. However, as noted above, the behavioral-treatment literature provides little guidance on how to select a specific treatment from the many options of treatments available for a specific functional class of problem behavior. Thus, applying the methods and concepts from the precision medicine paradigm to behavior analysis could lead to the identification of variables (in addition to the function of behavior) to guide selection of the best treatment for a given case. This approach requires identifying variables associated with positive (and negative) response to treatment and demonstrating that those relations have generality to the extent that we can make predictions about future outcomes with other individuals. Given that the prediction and control of behavior is a fundamental goal of the science of behavior (Skinner, 1953), exploring the application of methods used to identify PBioMs to the behavioral treatment of problem behavior warrants conceptual consideration and empirical evaluation.

Whereas PBioMs are defined generally as objective biological measures that predict response to treatment, *predictive behavioral markers* (PBM) can be defined as objective behavioral measures that predict response to treatment (Hagopian, Rooker, Zarcone, Bonner, & Arevalo, 2017).¹ Here, we illustrate that the methods used to identify, quantify, and validate PBioMs are conceptually compatible with applied behavior analysis and therefore, can be extended to PBMs. Briefly, these methods are essentially conditional probability analyses that quantify the probability of a favorable outcome, given the presence of a particular PBioM (or PBM) of interest. Though these methods require analysis of multiple cases to identify relations that have *generality* across participants, datasets are not aggregated in a manner that is incompatible with the analysis of individual behavior—a fundamental dimension of applied behavior analysis (Baer, Wolf, & Risley, 1968; Johnston & Pennypacker, 2009). Rather, these methods involve post hoc analysis of treatment data, in

¹Predictive behavioral markers are different from “predictive behavioral risk markers,” or “risk factors” which are behavioral measures found to co-occur with a clinical problem, but do not predict response to treatment. Although these are thought to indicate risk, causality can-not be determined from these correlational studies (see Davies & Oliver, 2016).

which individuals are categorized as either responders or nonresponders to treatment, followed by a process for identifying the variables that distinguish these responders from nonresponders.

Overview of methods used to identify and quantify PBioMs

We describe the quantitative methods used to identify PBioMs (or PBMs) in detail in the Methods section, but the general process is briefly summarized here. For the purposes of this discussion, the process of identifying a PBioM retrospectively can broadly be characterized as involving three general phases. First, the “candidate” PBioMs are identified through retrospective analysis of clinical-trial data; wherein individuals classified as responders or nonresponders to the treatment are examined further to determine if there are any biological variables that appear to distinguish these groups (La Thangue & Kerr, 2011). For treatment of medical conditions, objective benchmarks defining responders and nonresponders are disease-specific, but based on clinically meaningful outcomes (e.g., survival rate) rather than statistically significant improvements. Second, if a variable appears to distinguish responders from nonresponders, researchers then subject this candidate PBioM to analyses designed to quantify its accuracy (essentially examining the conditional probability of successful and unsuccessful outcomes given the presence of the PBioM). Researchers also can examine candidate PBioMs prospectively based on a priori hypotheses relating the mechanisms of disease and drug action, which are then tested in a prospective treatment study (prospective methods are described further in the Discussion). Finally, if the quantitative analysis of the PBioM demonstrates it is sufficiently accurate in discriminating responders from nonresponders, then a replication study is conducted to examine its generality beyond the original sample, and thus demonstrate its predictive utility.

Because the goal of this approach is to identify how well variables predict response to treatment, any retrospective analysis of treatment data must draw from a sample that includes both responders and nonresponders to treatment. Furthermore, the candidate predictor variables of interest must also vary within the sample if a relation is to be identified. PBioMs are sometimes identified retrospectively from data drawn from randomized clinical trials using group designs, but randomized clinical trials are rarely used in applied behavior analysis. However, large-scale analyses of experimentally controlled treatment outcomes across a series of consecutively encountered cases (in which all cases are included regardless of outcomes) are being used with increasing frequency in applied behavior analytic analysis studies (e.g., Greer, Fisher, Saini, Owen, & Jones, 2016; Hagopian, Rooker, & Zarcone, 2015; Jessel, Ingvarsson, Metras, Kirk, & Whipple, 2018; Scheithauer, Cariveau, Call, Ormand, & Clark, 2016). Because these designs include all cases encountered regardless of outcome, they minimize selection bias favoring a particular outcome. If the sample includes both responders and nonresponders, these studies provide a good source of data for the initial identification of candidate PBMs. Other sources of data might include large-n uncontrolled consecutive case-series studies (National Cancer Institute, n.d.), and clinical replication series studies (Hersen & Barlow, 1976); but because these methods do not require that each case undergoes a controlled experimental analysis, conclusions based on an analysis of those types of datasets would be tentative. Finally, multiple data sets can also be sourced from multiple small-n published studies reporting on

experimentally controlled, individual treatment analyses, but only if those studies include responders and nonresponders to treatment and are otherwise thought to comprise a representative sample of a given condition or population.

To demonstrate the applicability of the established methods for quantifying the accuracy of PBM to applied behavior analysis, we examined two candidate PBMs for ASIB: the level of differentiation of ASIB in the functional analysis, and subtype classification. We obtained the data for the current analysis by combining data from a consecutive controlled case series study (Hagopian et al., 2015), and with data from a quantitative literature review of all published cases of ASIB (Hagopian et al., 2017). We provide a brief review of relevant findings on ASIB below.

AUTOMATICALLY REINFORCED SELF-INJURIOUS BEHAVIOR

Self-injurious behavior (SIB) is usually maintained by social contingencies, but in approximately 20% to 25% of cases, functional-analysis results indicate SIB occurs independent of social contingencies. The term automatic reinforcement is used to describe this functional class because it is assumed the behavior produces its own reinforcement (Vaughan & Michael, 1982). Recent research has identified subtypes of ASIB based on unique patterns of responding in the functional analysis and the presence of self-restraint, a behavior that interferes with or is incompatible with SIB (Hagopian et al., 2015). The distinguishing feature of Subtype-1 ASIB is that it occurs most frequently in the alone or ignore condition of the functional analysis and infrequently in the play condition (i.e., responding is highly differentiated across play and no-interaction conditions). Another pattern of responding indicative of ASIB involves SIB that is high and sometimes variable across all conditions (this pattern of high and undifferentiated responding is characteristic of Subtype-2 ASIB). ASIB co-occurring with self-restraint is the hallmark of Subtype-3 ASIB. Hagopian et al. selected these response features based on the premise that they might reflect the functional properties of ASIB that were unique to each subtype (Hagopian et al., 2015), and they codified these features into objective criteria that would permit reliable subclassification of ASIB (criteria for subtyping are summarized in the Methods in the current study and described in full detail in Hagopian et al., 2017).

Findings of the original subtyping study (Hagopian et al., 2015) most relevant to the current discussion are that subtypes of ASIB differ markedly in terms of their response to treatment. Subtype-1 ASIB was highly responsive to treatment involving reinforcement alone (at a level comparable to a group of individuals with socially maintained SIB), whereas Subtype-2 and -3 ASIB were highly resistant to treatment using reinforcement alone and often required additional treatment components including restraint or protective equipment. The investigators drew the participants in original study ($n = 39$ with ASIB, and $n = 13$ with socially maintained SIB) from a sample receiving treatment in a single facility; so, to test the generality of the subtyping model, they conducted a replication study. In the replication study, they included every identified published dataset of ASIB with sufficient data to permit subtyping (Hagopian et al., 2017; $n = 49$ with ASIB, and 13 with socially maintained SIB as a comparison group). The limited number of published cases with Subtype-3 ASIB with

sufficient data to permit subtyping and analysis of treatment outcomes precluded a complete analysis of that subtype.

The findings from the replication study corresponded with those of the original subtyping study with respect to Subtype-1 and -2 ASIB. The level of differentiation of ASIB in the functional analysis across the play and no-interaction conditions of the functional analysis correlated positively with response to treatment using reinforcement alone in both the original and replication study ($r = 0.61$ and 0.72 , both p values $< .0001$, respectively). Combining data from both studies showed that reinforcement alone effectively reduced SIB by at least 80% in 19 of 23 cases (82.6%) with Subtype-1 ASIB (a level of responsiveness comparable to socially maintained SIB), but in only 1 of 14 cases (7.2%) with Subtype-2 ASIB.

Based on the consistency of the findings of the two subtyping studies, Hagopian et al. (2017) suggested that ASIB should no longer be considered a single category—and that the identified subtypes may likely have different underlying mechanisms. The assessment and treatment findings show that the rate of Subtype-1 ASIB varies inversely with the level of stimulation in the environment, suggesting treatment involving reinforcement may simply provide alternative sources of sensory stimulation that successfully compete with the sensory stimulation produced by SIB. The insensitivity of Subtype-2 ASIB to changes in the environment (low differentiation in the functional analysis and limited or no response to treatment using reinforcement) suggests alternative sources of reinforcement applied in those analyses did not compete with the presumed biological process or processes that occasion and/or maintains Subtype-2 ASIB. The authors suggested level of differentiation of ASIB in the functional analysis could be viewed as an index of sensitivity of SIB to disruption by alternative reinforcement, which is also evident in the context of treatment. Hagopian et al. (2017) also suggested that the level of differentiation of ASIB in the functional analysis and subtype classification might function as PBMs, but awaited a formal quantitative analysis. Therefore, the purposes of the current article were to describe established methods used to identify PBioMs, apply these methods to formally quantify the accuracy of two candidate PBMs for ASIB, and suggest the potential value of this approach for behavior-analytic research and practice beyond the treatment of self-injury.

METHOD

Participants and Settings

For the current investigation, we included participants with either Subtype-1 or -2 ASIB described in the two subtyping studies conducted by Hagopian et al. (2015; 2017). We excluded participants with a third subtype (Subtype-3 ASIB, characterized by the presence of self-restraint) from the current study because we did not have a sufficient sample of cases of Subtype-3 ASIB with treatment data. Because our purpose for the current study was to examine how well variables predict treatment outcomes, we included only cases with treatment datasets available (total $n = 37$). Nineteen of the included datasets came from Hagopian et al. (2015), who reported on patients treated in a single setting, and 18 of the included datasets came from Hagopian et al. (2017), who reported on cases of ASIB drawn from the published literature. The cases from the 2017 study included individuals treated in:

schools ($n = 5$), residential programs and homes (4), outpatient programs (4), inpatient clinics (3), and unspecified settings (2). Table 1 summarizes demographic information for the sample of 37 included in the current study. We also included demographic data on the full sample of 78 datasets with Subtype-1 and -2 ASIB from the two Hagopian et al. studies as point of reference to show that the sample of 37 cases included in the current study provided a generally representative subsample of the larger group of patients with ASIB in terms of demographic variables.

Defining Response to Treatment

Methods used with PBioMs. Although treatment response can vary on a continuum, analysis of PBioMs requires distinguishing between participants who are either responders and nonresponders to treatment. Outcomes are based on objectively measurable and clinically meaningful direct measures of the disease (e.g., reduction in tumor size). The criteria that define what constitutes a responder and nonresponder are disease-specific and based on empirical findings related to what is considered to represent a level of improvement that confers clinically meaningful benefit. Additional measures that capture changes in quality of life (improvements as well as the burdens of treatment) that go beyond direct measures of the disease are used to better capture the impact of treatment on the person; however, these subjective measures are no substitute for a direct measure of disease improvement (Higginson & Carr, 2001).

Application in the current study.

We emulate methods used with PBioMs by examining an objective and direct outcome measure: reduction in SIB brought about by treatment. We distinguished responders from nonresponders based on whether treatment with a reinforcement-based intervention alone produced at least an 80% reduction in ASIB from baseline, a benchmark for successful treatment frequently used in behavior-analytic studies (e.g., Donaldson, Vollmer, Krous, Downs, & Berard, 2011; Greer et al., 2016). Because classification as responder and nonresponder is based on this direct measure of behavior change, findings should be interpreted only in terms of this specific outcome.

Continuous and Categorical Predictor Variables

Methods used with PBioMs.—A PBioM can be either a continuous variable (such as the level of a protein in the blood), or a categorical or ordinal variable (such as the presence of a certain genetic mutation). Each type of variable requires different quantitative methods to examine its predictive utility. A requirement of PBioMs is that they must be objectively measured using standardized procedures that permit replication across studies; and here we extend this requirement to PBMs.

Application in the current study.—We examined two candidate behavioral markers in the current study: (a) level of differentiation of ASIB across the play and no-interaction conditions of the functional analysis (a continuous variable); and (b) Subtype of ASIB (a binary categorical variable; i.e., Subtype-1 or -2 ASIB). Detailed descriptions of the functional-analysis methods, subtype classification, treatments, and treatment-evaluation methods are described in the studies from which we obtained these data (Hagopian et al.,

2015; 2017). We derived both subtypes from functional-analysis data using two standard conditions: the control (play) and no-interaction conditions (ignore or alone). During the play condition, the child is in a room with the therapist and preferred toys, during which time the therapist ignores problem behavior and interacts with the child every 30 s. During the no-interaction condition (alone or ignore), the investigators placed the child in a room without any play or leisure materials either alone, or with a therapist who monitored but did not interact with, or respond to the child. Below, we further describe how both candidate PBMs examined in the current study are consistent with the requirements of PBioMs in that they are objectively measured using procedures that would permit replication.

We calculated the first candidate PBM, the *level of differentiation* of ASIB in the functional analysis, by dividing the average rate of ASIB in the play condition by the average rate of ASIB in the no-interaction condition, subtracting the resulting quotient from 1, and then converting that value to a percentage. Thus, percentage differentiation refers to the reduction of ASIB in the play condition relative to the no-interaction condition. Proportional change in responding across conditions is a commonly used metric in behavior-analytic research (e.g., Asmus et al., 2004). The maximum level of differentiation of 100% indicates that the average rate of ASIB is 0 in the play condition, but greater than 0 in the no-interaction condition. A level of differentiation of 0% indicates the average rates of SIB are equal across no-interaction and play conditions. When the level of differentiation is a negative value, this indicates the average rate of ASIB is higher in the play condition relative to the no-interaction condition.

We derived the second candidate PBM, *subtype classification* (either Subtype-1 or -2 in the current study) based on formal, objective criteria. The criteria we used are based on general consensus in the field that the presence of automatically reinforced behavior is defined by one of two patterns of functional-analysis results: (a) highly differentiated responding in the alone or ignore condition relative to the play condition; and (b) high and variable rates of responding across all conditions (see Hagopian et al. 1997; Iwata, Dorsey, Slifer, Bauman, & Richman, 1982/1994; LeBlanc, Patel, & Carr, 2000; Vollmer, 1994). We described the objective criteria for subtyping ASIB in greater detail in the subtyping studies (Hagopian et al., 2015; 2017), which we derived from the criteria for interpretation of functional analysis data originally described by Hagopian et al. (1997) and further refined by Roane et al. (2013). Those criteria define differentiated responding in any test condition of the functional analysis based on the observation that half of the data points in a test condition (the alone or ignore condition in the case of ASIB) of the FA exceed 1 standard deviation above the mean of the play condition (along with other considerations related to trend, effect size, and variability, which are included in the criteria).

Determining Specificity and Sensitivity

Methods used with PBioMs.—Researchers analyze the predictive accuracy of PBioMs by examining their specificity and sensitivity. The *specificity* of a PBioM refers to the proportion of cases that the biomarker correctly identifies as nonresponders to treatment (“true negatives”). The *sensitivity* of a PBioM refers to the proportion of cases that the biomarker correctly identifies as responders to treatment (“true positives”). Because

specificity and sensitivity are inversely related, researchers must determine the relative importance of each. Thus, when the costs or risks (to the patient) of false positives and false negatives are roughly equal, it is desirable to maximize both specificity and sensitivity to the extent possible (without sacrificing one for the other). For some medical conditions, however, the relative costs of false positives and false negative may be unequal. For example, lower sensitivity and higher specificity may be desirable in cases in which over-identification could unnecessarily result in invasive, high-risk, or irreversible interventions (i.e., higher false positives would be undesirable). Conversely, for conditions in which failing to identify a problem results in rapid disease progression, it is desirable to have higher sensitivity, and “sacrifice” specificity to some degree (i.e., higher false positives would be preferred over false negatives).

Application in the current study.—Behavioral treatment using reinforcement is relatively benign and poses minimal short-term risks even if it is ineffective. Thus, for the current analysis, we sought to maximize both specificity and sensitivity to the extent possible (and thus not sacrifice one to improve the other).

Methods used with continuous PBioMs: Receiver Operating Characteristic (ROC) curves.—With continuous PBioMs, researchers use Receiver Operating Characteristic (ROC) curves to analyze the trade-off between specificity and sensitivity and quantify the predictive accuracy of the PBioM (Pepe, 2003). A ROC curve depicts levels of sensitivity (y -axis) and the inverse of specificity ($1 - \text{specificity}$; x -axis) for all obtained values of the PBioM, and thereby permits one to select the optimal “cutoff” point for the PBioM. Once that cutoff point is empirically determined, then it is used to make predictions about outcomes. Those individuals that exceed the cutoff, are characterized as “testing positive” for the PBioM, and are predicted to have a favorable response to treatment. Most statistical analysis software programs can calculate ROC curves based on the following formulas. The value of the candidate PBioM is denoted as Y , with higher values of Y associated with positive treatment outcome ($D = 1$). Using a cutoff score, c , we calculate the proportion of individuals that would be considered to have positive outcome (responders) based on the value of the candidate PBioM as follows: positive, if $Y \geq c$; and negative, if $Y < c$. Then for each c , *true positive fraction* (TPF ; which is the same as sensitivity) can be defined as the proportion of cases labeled as responders by the candidate PBioM among the individuals with positive treatment outcome (i.e., responders, $D = 1$), or as conditional probability:

$$TPFc = P(Y \geq c | D = 1)$$

Analogously, false positive fraction (FPP) can be defined as the proportion of individuals labeled positive by the candidate PBM who did not have a positive treatment outcome (i.e., nonresponders, $D = 0$), or as conditional probability. FPP is equivalent to $1 - \text{specificity}$, and is calculated as using this formula:

$$FPPc = P(Y \geq c | D = 0)$$

As c varies, the threshold for positive treatment outcome based on the candidate PBioM value, one obtains the entire set of TPF and FPF pairs for each c (i.e., for every obtained level of differentiation in the sample). These pairs of TPF and FPF can be plotted to produce the ROC curve, in which TPF is plotted on the y -axis and FPF is plotted on the x -axis. As the value of c increases, the points (defined by the pair of TPF and FPF) move from the lowest left corner of the plot to the upper right corner. Because both TPF and FPF are almost always less than or equal to 1, the area *under the curve* (AUC), which refers to the accuracy of prediction, of the plot of the ROC Curve is a square with an area of 1 (see Figure 1).

In case of the perfect classification, the PBioM perfectly classifies cases as responders versus nonresponders, such that $TPF = 1$ and $FPF = 0$ (line A in Figure 1). In this case, the ROC curve moves towards the upper left corner as the AUC gets closer to 1 (meaning the predictor approaches 100% accuracy). The circled data point on line A represents the value of the PBioM at which the maximum combination of specificity and sensitivity is attained (in the case of perfect classification, sensitivity and specificity are both 1.0). On the other hand, if the candidate PBM is useless for determining the treatment response status, at every c , TPF_c will be the same as FPF_c . In this case, the ROC curve will be the 45° line (line C in Figure 1), with $AUC = 0.5$ (i.e., the predictor is 50% accurate; no better than chance at predicting a positive or negative outcome). By evaluating the AUC, one will be able to determine whether a candidate PBM is useful in predicting treatment response (i.e., its accuracy). This is done by assessing how far numerically the AUC is away from 0.5 (at which there is no relation between the candidate PBM and treatment outcome) and how close it is to 1 (at which there is a perfect relation).

Line B of Figure 1 represents a relation between a PBM and treatment outcome that is intermediate between perfect relation and no relation. As a general rule, AUC values above 0.7 indicate acceptable discrimination (>70% accuracy), and AUC values above 0.8 (> 80% accuracy) indicate excellent discrimination between patients who will or will not respond to the treatment (Hosmer & Lemeshow, 2000). The circled data point on line B represents the value of the PBioM at which the maximum combination of specificity and sensitivity is attained. To select an optimal cutoff for a continuous PBioM, one might choose the optimal values for TPF and FPF for a particular situation (i.e., disease) or even look at the cost associated with various levels of these indices (Pepe, 2003).

Application in the current study.—The level of differentiation of ASIB (proportional responding across the play and no-interaction condition of the functional analysis) is a continuous variable. To examine this as a candidate PBM, we conducted an ROC curve analysis using methods described above. In essence the TPF (sensitivity) and FPF ($1 - \text{specificity}$) are calculated for each obtained value of the level of differentiation (c) for each case. The point at which sensitivity and specificity are maximized represents the optimal cutoff point for the PBM where responders and nonresponders are most accurately distinguished. In the current study, the analysis was performed using STATA-14 software (StataCorp, 2015), but it is possible to perform the calculations by hand by calculating TPF and FPF for each obtained value of level of differentiation.

Methods used with categorical (or binary) PBioMs.—*TPF* and *FPF* are also applicable to categorical or binary PBioMs, such as the presence ($Y = 1$) or absence ($Y = 0$) of a genetic mutation that predicts response to treatment. Using the same notation as above, *TPF* (which is the same as *sensitivity*), is the proportion of individuals with positive outcomes who are correctly labeled as such by the binary PBioM, in which $P(Y = 1 | D = 1)$. Similarly, $1 - FPF$, or *specificity*, is the proportion of individuals with negative outcome who are correctly identified as such by the PBioM, in which $P(Y = 0 | D = 0)$. Here, “testing positive” for the predictor refers to the presence of the characteristic thought to be associated with a favorable treatment response ($Y = 1$), and “testing negative” refers to the absence of that characteristic ($Y = 0$). Calculations can be performed using the following formulas:

$$\text{Sensitivity} = \frac{\text{n of cases for which } Y = 1 \text{ that were responders}}{\text{total n of responders}}$$

$$\text{Specificity} = \frac{\text{n of cases for which } Y = 0 \text{ that were nonresponders}}{\text{total n of nonresponders}}$$

Application in the current study.—Subtype classification is a categorical variable, and in the present study a binary categorical variable, as SIB can be classified as either Subtype 1 or Subtype 2. Using a positive treatment response as the positive outcome ($D = 1$), and Subtype-1 ASIB ($Y = 1$) as the predictor of treatment response, sensitivity is the proportion of cases with Subtype-1 ASIB among those who respond to treatment. Specificity is the proportion of cases without Subtype-1 ASIB (i.e., they have Subtype-2 ASIB; $Y = 0$), among the nonresponders. STATA 14 (StataCorp, 2015) was used to conduct the analyses described in the current study; however, these calculations can be performed easily by hand using the following formulas:

$$\text{Sensitivity} = \frac{\text{n of cases with Subtype-1 ASIB that were responders}}{\text{total n of responders}}$$

$$\text{Specificity} = \frac{\text{n of cases with Subtype-2 ASIB that were nonresponders}}{\text{total n of nonresponders}}$$

Metrics for Making Clinical Predictions

Methods used with PBioMs.—Sensitivity and specificity are useful metrics for describing and classifying the accuracy of a PBioM. However, in practice, a more important question is one of prediction: that is, how well one can predict the treatment outcome if the value of the PBioM is known? Therefore, instead of being interested in the proportion of individuals who test positive for PBioM among those who respond to treatment (i.e., sensitivity), a more important quantity is *the proportion of individuals who are responders among those that “test positive” for the PBioM*. This quantity, the *positive predictive value* (PPV), is the conditional probability of having a good outcome given the presence of the PBioM that is thought to predict a favorable response to treatment. *PPV* is calculated using

this formula ($Y = 1$ indicates that the individual “tested positive” for the PBioM thought to predict a good response to treatment):

$$PPV = \frac{\text{n of cases for which } Y = 1 \text{ that were responders}}{\text{total n of cases for which } Y = 1}$$

Similarly, *negative predictive value (NPV)*, is the conditional probability of not having a good outcome, given the absence of the PBioM that is thought to predict favorable response to treatment, and is calculated using this formula ($Y = 0$ indicates the individual “tested negative” for the PBioM thought to predict good response to treatment):

$$NPV = \frac{\text{n of cases for which } Y = 0 \text{ that were nonresponders}}{\text{total n of cases for which } Y = 0}$$

Application in the current study.—Specific to our PBM analysis, PPV is the conditional probability of a good outcome, given classification as Subtype-1 ASIB, and is calculated using the following formula:

$$PPV = \frac{\text{n of Subtype-1 cases that were responders}}{\text{total n of Subtype-1 cases}}$$

Similarly, *negative predictive value (NPV)*, the conditional probability of not having a good outcome, given not being classified as Subtype-1 ASIB (or being classified as Subtype-2 ASIB), is calculated using this formula:

$$NPV = \frac{\text{n of Subtype-2 cases that were nonresponders}}{\text{total n of Subtype-2 cases}}$$

RESULTS

Level of Differentiation of ASB as a PBM

Figure 2 shows the obtained ROC curve using the level of differentiation of ASIB in the functional analysis (proportional responding across the play and no-interaction conditions) as a PBM for treatment outcome. The AUC for this PBM was 0.87 (95% confidence interval; CI from 71% to 95%), a value considered to represent a *good to excellent* prediction of treatment outcomes. This indicates we correctly classified 87% of participants (32 of the 37) as either responders or nonresponders with the derived cutoff score of 63.7% differentiation (sensitivity = 0.91; and specificity = 0.82; See Table 2). Moreover, this cutoff score produced a PPV of 85.7% (CI from 64% to 97%; the wide confidence interval is due to the small sample size), which means that among all individuals who “test positive” for high differentiation (meeting or exceeding the cutoff of 63.7% level of differentiation), 85.7% are expected to be responders to treatment using reinforcement alone. The PPV statistic essentially quantifies the extent to which we can predict and control ASIB when the level of differentiation in the FA exceeds 63.7%. This cutoff score produced a negative predictive value (NPV) of 87.5%, indicating that when the level of differentiation of ASIB is below 63.7% (“test negative” for high differentiation), 87.5% would not be expected to have

a favorable response to reinforcement alone (thus, only 12.5% of these cases would be expected to have a favorable outcome).

Analysis of Subtype Classification as a PBM for Automatically Reinforced SIB

Across the entire sample, 20 of the 37 participants (54.1%) showed a positive response to reinforcement-based treatment (i.e., 80% or greater reduction). However, response to reinforcement-based treatment differed markedly across subtypes: 82.6% of cases (18 of 21) with Subtype-1 ASIB responded favorably to a reinforcement-based treatment, but only 7.1% (1 of 14) of those with Subtype-2 responded favorably (see Table 3). Thus, the PPV for Subtype-1 was 82.6% (95% CI from 61% to 95%), indicating that there is 82.6% chance that an individual classified with Subtype-1 ASIB would respond to treatment using reinforcement alone (see Table 3). This PPV essentially quantifies the extent to which we can predict and control Subtype-1 ASIB. The NPV was 92.9% in this sample (95% CI from 66% to 100%). This means that among individuals with Subtype-2 ASIB, 92.9% would be predicted to not have a successful treatment response (i.e., nonresponders) to reinforcement alone. Although this NPV indicates we can accurately predict that Subtype-2 ASIB is unlikely to respond to reinforcement alone, it also quantifies our limited ability to control Subtype-2 ASIB using reinforcement alone.

Level of differentiation and subtype classification are both determined by assessing responding in the no-interaction and play conditions of the functional analysis, and thus are closely related but slightly different metrics. Level of differentiation represents the proportional rate of ASIB in the play condition relative to the no-interaction conditions of the functional analysis; whereas subtype classification represents one of two patterns of responding in the functional analysis characterized by either sensitivity or insensitivity to disruption by alternative reinforcement. The ROC curve derived cutoff of 63.7% differentiation level corresponds closely to the intermediate point of differentiation that distinguishes cases in which ASIB was classified as either Subtype-1 or Subtype-2; which further supports the proposed categorical subtyping model.

Figure 3 depicts each participant's data as a coordinate of the relation between level of differentiation in the FA and percentage reduction with treatment using reinforcement alone. Subtype classification is indicated by the symbol, the vertical dashed line indicates the ROC-derived, level-of-differentiation PBM cutoff point (63.7% differentiation), and horizontal line indicates the effective treatment benchmark distinguishing responders from nonresponders (80% reduction from baseline). The four quadrants represent: false-positives, true-positives, true negatives, and false negatives. Using the level of differentiation as the PBM, we correctly classified 32 of the 37 cases as either responders (18 true-positives) or nonresponders (14 true-negatives). Of the five cases that we classified incorrectly, (a) two were false-negatives, wherein treatment was successful despite differentiation being below the cutoff; and (b) three were false-positives, wherein the level of differentiation exceeded the cutoff, but treatment was not successful.

DISCUSSION

The goals of prediction and control are not exclusive to the science of behavior but are important in many applied fields, including precision medicine. This article illustrates how methods and concepts used in precision medicine can be extended to applied behavior analysis. Specifically, we demonstrated how the established methods used to identify predictive biomarkers (PBioms) can be applied to identify predictive behavioral markers (PBMs). Following a review of the implications of these findings with respect to ASIB, we discuss the potential of these methods for the field of applied behavior analysis.

Discussion of Findings Related to ASIB

The results of the analyses performed in the current study indicate that the level of differentiation of ASIB in the functional analysis, and ASIB subtype classification (as Subtype 1) each can be characterized as *good to excellent* PBMs for response to treatment using reinforcement alone. However, replication of these results using a larger cohort of individuals would be important to more fully validate the outcomes. By quantifying the accuracy of these two PBMs, we demonstrate that other dimensions of problem behavior (in conjunction with the functional class; Iwata, Pace, Cowdery, & Miltenberger, 1994) may accurately predict treatment outcomes. That is, although all the participants had SIB maintained by automatic reinforcement, the two subtypes of ASIB differed markedly in their responsiveness to reinforcement-based treatment based on another dimension of behavior. Although these findings describe the observed relation between these PBMs and treatment outcomes for the cases included in this analysis, the extent to which these cases are representative of the broader population is difficult to determine. We obtained roughly half of the data sets from a single specialized treatment facility (Hagopian et al. 2015) and the remainder from the published literature (Hagopian et al., 2017) that included individuals treated in multiple settings by multiple clinical researchers. It is possible that referral bias affected the results of the first study, and publication bias affected the results of the second study. However, those two studies produced highly concordant results, which help to mitigate concerns about these potential threats to external validity. The fact that both samples included so many cases showing a favorable response for Subtype-1 ASIB and much poorer and more varied outcomes for Subtype 2 makes it difficult to see how biases could have substantively affected the results. Nevertheless, the results describe observed relations between the PBMs and treatment outcomes for these cases, and represent only probable outcomes that may be obtained with future cases with similar PBM values.

The ROC curve calculated for the level of differentiation of ASIB in the functional analysis, derived the optimal cutoff (i.e., 63.7% differentiation) that maximized the area under the curve (AUC) resulting in 87% predictive accuracy. For cases with ASIB that “test positive” for either highly differentiated ASIB in the FA (exceeding the cutoff), or those that “test positive” as having Subtype-1 ASIB, the obtained positive predictive values (PPV) of 85.7% and 82.6%, respectively, represent the proportion of cases for which reinforcement alone is likely to be effective. These values provide one way to quantify the extent to which we can predict and control Subtype-1 ASIB (through treatment). For those who “test negative” for either high differentiation, or for Subtype-1 (i.e., are classified as having Subtype-2 ASIB),

the obtained negative predictive values (92.9% and 87.5%) represent the proportion of cases for which reinforcement is not effective. These values quantify the extent to which we can predict negative outcomes; but they also illustrate our limited ability to control Subtype-2 ASIB (i.e., effectively treat ASIB with reinforcement alone). As noted previously, differences across subtypes suggests that ASIB probably should not be considered a single category, and highlights the need to better understand and develop more effective interventions for Subtype-2 ASIB. Although Subtype-1 ASIB appears generally homogenous, the observed heterogeneity of Subtype-2 in terms of the variation in the level of differentiation in the functional analysis and the widely diverse response to treatment raises questions about whether Subtype-2 ASIB represents a single subcategory (Hagopian et al., 2017; Hagopian & Frank-Crawford, 2017).

Translating Precision-Medicine Methods and Concepts to Applied Behavior Analysis

Compatibility with applied behavior analysis.—The methods used to identify PBioMs and to evaluate their accuracy are fundamentally compatible with key dimensions of applied behavior analysis (Baer, Wolf, & Risley, 1968); and thus, are readily applicable to PBMs. First, identifying PBMs is an *applied* endeavor in that its primary aim is to guide the selection of the most *effective* treatment for a given individual (though the discovery of PBMs could also have scientific value as they may elucidate the mechanisms involved in the maintenance of problem behavior and treatment action). Second, the approaches described in the current study are behavioral, technological, and analytic in that both the predictor and outcome measures represent intrasubject changes in behavior identified in the context of controlled experimental analyses of behavior. In the current study, the predictor and outcome variables represented within-subject proportional changes in SIB across assessment and treatment conditions identified using a single-subject experimental design. This approach does not involve combining or averaging data from individual cases in a way that obscures individual behavior change but rather compiles data from multiple cases to identify relations that have *generality* across individuals, but are not evident at the level of the individual participant. Finally, by determining the conditional probability of a future successful outcome (i.e., the Positive Predictive Value) when an individual “tests positive” for a PBM, we essentially quantify the prediction and control of behavior—a fundamental goal of the science of behavior (Watson, 1913).

Identifying PBMs builds upon the function-based approach to treatment of problem behavior that has guided practice and research in applied behavior analysis for decades. Research shows that behavioral interventions informed by functional analysis findings are more effective than those based on descriptive methods (Didden, Korzilius, van Oorsouw, & Sturmey, 2006; Hurl, Wightman, Haynes, & Virues-Ortega, 2016), but retrospective studies examining differences between responders and nonresponders to behavioral treatment within specific functional classes have yet to be conducted. Prospective studies comparing treatments that are indicated and contraindicated based on the function of problem behavior are limited to a just few. Iwata et al. (1994) demonstrated that extinction procedures that matched the function of behavior effectively reduced the target behavior whereas mismatched extinction procedures did not. A handful of other studies have produced similar results (Brown et al., 2000; Kern, Delaney, Hilt, Bailin, & Elliot, 2002; Kuhn, DeLeon,

Fisher, & Wilke, 1999; Richman, Wacker, Asmus, & Casey, 1998). Although predictive accuracy of the function of problem behavior has not been formally quantified using the methods described in the current study, we can characterize it as PBM in light of the extensive body of evidence supporting function-based interventions and the status of this approach as reflecting best practice. Moving forward, methods described in this study derived from the field of precision medicine can provide a blueprint for research aimed at identifying and quantifying additional PBMs.

Potential Benefits of Identifying PBMs

Although identifying effective interventions is critical to advancing practice, identifying variables associated with both success and failure can advance knowledge in a different way. If advances achieved with PBioMs are any indication, identifying PBMs has the potential to increase efficiency and improve care, inform research, elucidate mechanisms of problem behavior and treatment action, and help define the indications and limitations of assessment and treatment procedures.

Improving efficiency.—First, identifying additional PBMs would enable practitioners to select the most effective intervention for a specific problem sooner rather than later, saving time and financial resources. The benefit of efficiency has been realized in medicine, and is particularly important for diseases that can advance quickly and become more difficult to treat over time. For the treatment of severe problem behavior that can produce injury, permanent disfigurement, loss of function, or restrictive placement, applying the most effective intervention sooner rather than later could have a profound impact on the individual and his or her family. Moreover, knowing that particular types of behavior (e.g., Subtype-2 ASIB) are highly resistant to treatment using reinforcement alone and will likely require protective equipment, or more restrictive procedures, such as punishment or restraint, can inform long-term clinical planning for individual cases. For skill acquisition programming, targets and goals change often, so even small improvements in treatment efficiency accumulated across multiple programs and over the span of a few years could have a meaningful impact. This might be particularly important for young children with autism, considering what is known about the benefits of early intervention (Eldevik et al. 2009), and in light of the increasing prevalence of autism and rising financial costs for quality behavior-analytic interventions.

Informing research and advancing knowledge.—Second, identifying PBMs has the potential to inform research aimed at understanding problem behavior, identifying sources of heterogeneity including subtypes, and determining the mechanisms of treatment action. For example, the identification of treatment-responsive and treatment-resistant subtypes of ASIB raises questions about their underlying mechanisms and the mechanisms by which treatments engender behavior change. The way in which the rate of Subtype-1 ASIB varies inversely with the availability of alternative reinforcement is highly characteristic of a reinforced response as predicted by matching in single-alternative arrangement (Reed & Kaplan, 2011). In contrast, the resistance of Subtype-2 ASIB to reinforcement might suggest it may be a reinforced operant that produces highly potent biological reinforcement, greatly limiting the impact of alternative reinforcement (using currently available methods). It is

also possible that this is not an operant response, but the product of a repetitive movement disorder, or related to sensory dysfunction (Hagopian & Frank-Crawford, 2017). Thus, the identification of additional PBMs would have the potential to inform research aimed at understanding *why* the PBM predicts response to treatment, which could advance knowledge about the problem itself and the mechanisms of treatment action, and direct and inform research designed to make existing treatments more effective.

Defining the optimal uses of clinical procedures.—Identifying PBMs also has the potential to help us build a body of evidence to define the indications and limitations of specific treatments. Most behavioral treatment studies are designed to examine a single treatment's effectiveness, whereas a minority of behavioral treatment studies compare multiple treatments to determine if one is superior to the other across cases (e.g., Vollmer, Iwata, Zarcone, Smith, & Mazaleski, 1993). Although this latter type of research is informative, broad comparisons of the effectiveness of different interventions generally do not evaluate the possibility that effectiveness may be conditional rather than absolute. That is, this approach to examining variables associated with success and failure could be used to determine the *conditional effectiveness* of treatments, broadly defined as the differential effectiveness of a treatment across circumstances (including the presence of a certain PBM). Thus, if future researchers identify additional PBMs, the accumulated evidence could also help define the conditional effectiveness of multiple treatments, and thus begin to define the indications (and contraindications) for each treatment. Thus, if several studies found Treatment A to be highly effective for certain functional classes of behavior (e.g., social positive reinforcement, and automatic reinforcement), but not for other classes (e.g., social negative reinforcement) for which Treatment B is most effective, then we could define the optimal uses for each treatment.

The same methods used to determine the conditional effectiveness of treatments could also be extended to other clinical procedures, including assessment procedures. Whether in the context of assessment or treatment, any information that would inform the selection of the best clinical procedure likely to be effective, and forgo attempting those that are unlikely to be effective could greatly improve efficiency and care. For example, the interview-informed synthesized contingency analysis (IISCA) has been demonstrated to be effective (Ghaemmaghami, Hanley, & Jessel, 2016; Jessel, Hanley, & Ghaemmaghami, 2016; Jessel et al., 2018; Slaton, Hanley, & Raftery, 2017) and has been directly compared to analog functional analysis (Fisher, Greer, Romani, Zangrillo, & Owen, 2016; Slaton et al., 2017). Though direct comparisons of procedures are useful, another approach would employ the methods and concepts described in the current study to help identify variables (e.g., participant characteristics, response characteristics, etc.) that discriminate responders from nonresponders to each approach. Determining the conditional effectiveness of these assessment methods would identify the indications for conducting the IISCA (which can be completed more quickly), versus a more intensive traditional functional analysis. Simply recognizing that the effectiveness of clinical procedures can be conditional rather than absolute, can focus efforts directed at determining the circumstances under which each clinical procedure is maximally effective.

Potential Classes of PBMs

As noted, precision medicine informs treatment selection for the individual patient based on: (a) an understanding of the causal mechanisms of disease processes and (b) the mechanisms of treatment action. Thus, PBioMs include genetic mutations that are related to the disease (McDonald et al., 2017), as well as genetic mutations that affect whether a drug can be properly metabolized by the body and thus produce the intended effect (pharmacogenomics; Scott, 2011). We propose two potential classes of PBMs that are analogous to these classes of PBioMs. We have defined a PBM as an objective behavioral measure that predicts response to treatment, so any candidate PBM *must be measured objectively through direct behavioral observation*.² The examples of possible candidate PBMs within each of the two classes discussed below are provided merely for illustration purposes, and have little or no empirical basis at the current time.

Target behavior PBMs.—The first class of PBMs analogous to disease-related PBioMs might include variables related to any dimension of the *target behavior* that could be shown to predict response to treatment. These *target behavior PBMs* would include dimensions of behaviors targeted for reduction, but could also include behaviors targeted for increase in skill acquisition programming. The operant function of problem behavior, the level of differentiation of ASIB in the functional analysis, and subtype of ASIB are examples of target behavior PBMs. As noted above, it seems that other likely candidate PBMs would be those that measure some dimension of problem behavior related to its controlling variables and the likely mechanism of treatment action. This could include measures of how the problem behavior changes in a changing environment (e.g., its resistance to extinction, or sensitivity to other types of disruptors), the degree to which responding continues in absence of the establishing operation in the escape or tangible conditions, and the target response's relation to other behaviors (e.g., its membership in a hierarchical, functional response class; Lalli, Mace, Wohn, & Livezey, 1995).

Behavioral capacity PBMs.—The second class of PBMs that might be worth exploring is analogous to PBioMs that affect therapeutic metabolism of drugs. These *behavioral capacity PBMs* are especially relevant for individuals with intellectual and developmental disabilities who may vary in terms of whether they possess the pre-requisite skills necessary for certain classes of behavioral interventions to impact performance and learning. Behavioral interventions that involve simple contingency shaping using primary reinforcers likely can affect behavior change in most individuals, but some individuals with more severe intellectual disabilities may be less responsive to increasingly complex arrangements involving conditional discriminations, conditioned reinforcement, timing, verbal mediation, rule following, stimulus equivalence, etc. If, for example, an individual is unable reliably to make conditional discriminations (or would require extensively more learning trials), an intervention including such a component that might otherwise be effective, may be less effective or impractical to implement for that individual relative to others. Other types of skills deficits that could impact effectiveness of certain types of treatments might include oversensitivity to temporal discounting (Critchfield & Kollins, 2001), and repetitive or

²Indirect measures of behavior such as scores on rating scales would not be considered PBMs.

invariant responding (Miller & Neuringer, 2000). These behavioral capacities could apply to treatments aimed at problem behavior reduction as well as skill acquisition. It seems plausible that the more pronounced and generalized one's performance deficits are, the more likely they could alter the effectiveness of specific behavioral interventions that rely more heavily on those skills.

Use of special assays.—We identified the two PBMs in the current study using functional analysis data, which are commonly obtained in the course of providing clinical care in many settings. Although functional analysis data could be the source for additional candidate PBMs, identifying PBMs related to other dimensions of the target behavior or behavioral capacities, may be possible only by conducting special assays or assessments designed to explicitly examine a dimension of interest of the target behavior or behavioral capacity thought to be a candidate PBM (or established as a PBM). This would be analogous to analyzing biological samples to obtain a measure of a candidate or established PBioM (e.g., using a blood sample to identify the level of a protein). For example, if one wished to examine a problem behavior's resistance to extinction as a candidate PBM, this would require a special assay because a typical functional analysis would not expose the target behavior to extinction (or would do so only in an extremely limited manner for a limited number of cases; i.e., using an ignore condition for attention-maintained behavior). Such an assay may be similar to a response-class hierarchy analysis (Lalli et al., 1995; Richman, Wacker, Asmus, Casey, & Andelman, 1999), or a precursor analysis including a conditional probability analysis (Fritz, Iwata, Hammond, & Bloom, 2013). However, evaluating any candidate PBM would require a sample that includes enough cases so that: a) the candidate PBM varies across a range of values, and b) response to treatment varies across a range of outcomes (and includes both responders and nonresponders). The sample size needed would depend on how accurately the candidate PBM predicted response to treatment.

Special assays might also be needed to systematically quantify many candidate PBMs related to behavioral capacities such as temporal discounting, as these are not routinely measured in the course of clinical care or educational programming. For example, Vollmer, Borrero, Lalli, and Daniel (1999) conducted an analysis to test for impulsivity, and showed that for two highly impulsive children, signaled delay schedules were more effective than unsignaled delay schedules (the study did not include nonimpulsive children to test if they needed signaled delays as well). A behavior analyst interested in examining impulsivity as a candidate PBM could employ a special assay designed to test impulsivity like that described by Vollmer et al. (1999) or Neef et al. (2005). The impulsivity assay would be obtained with the entire sample of cases (including impulsive and nonimpulsive cases), and then treatment with and without a component thought to address or overcome impulsive tendencies (i.e., the use of signaled delays) would be applied to examine whether the candidate PBM (i.e., sensitivity to reinforcer immediacy) accurately predicts response to treatment designed to address a behavioral capacity deficit.

Prospective Methods

Retrospective methods, such as those described in the current study, involve post-hoc analysis of treatment outcome data. This approach retrospectively examines whether responders and nonresponders differed on some variable, and then subjects that candidate PBM to analyses examining conditional probabilities of outcomes. By contrast, prospective methods are designed to test a priori hypotheses about the predictive value of a candidate PBM by comparing outcomes of a treatment tailored to the candidate PBM (i.e., a “matched” treatment) in one of several ways. One type of comparison would involve analysis of outcomes across cases that “test positive” for the PBM and are expected to be responders, relative to those who “test negative” for the candidate PBM and are expected to be nonresponders. Other types of comparisons would examine outcomes obtained with treatment that is “matched” to the candidate PBM relative to: (a) a “standard of care” treatment (i.e., the pre-dominantly used treatment), or (b) to a contraindicated treatment or “mismatched” treatment that would be hypothesized to not be as effective for those who test positive for the PBM. Prospective analysis of many candidate PBMs would likely require the use of special assays, developed specifically to examine a candidate PBM hypothesized to have predictive value for a treatment tailored to it. Clinical staff applying the intervention should be blind to the outcome of special assay to assess the PBM, in order to minimize biases that might affect treatment application. Such a prospective analysis of PBMs would mirror the type of analysis performed by Iwata and colleagues (1994) involving comparisons of extinction treatments that were matched and mismatched to the function of problem behavior. Similarly, analyses of the EGFR mutation in non-small-cell lung cancer as a PBioM have included comparisons of the effects of EGFR inhibitors (the “matched” treatment) to those testing positive and negative for the mutation, and comparisons of EGFR inhibitors to standard chemotherapy (a “standard of care” treatment) for those who test positive (Okimoto & Bivona, 2014). Whether using retrospective or prospective methods, examining the accuracy of a PBM requires the sample be composed of both responders and nonresponders to treatment, and that the PBM varies across the sample. Although an exploratory study to develop special assays or conduct preliminary hypothesis testing could be done with a relatively small sample, a larger sample would be needed to test the accuracy of a PBM.

Conclusions

Precision medicine represents a paradigm shift in medicine that brings it into clearer conceptual alignment with applied behavior analysis. This approach relies on biomarkers to diagnose and classify diseases, predict their course, and guide treatment selection. The current study (a) describes the methods used to quantify accuracy of PBioMs, (b) suggests they are compatible with fundamental tenets of behavior analysis, and (c) provides a demonstration of their application to identify PBMs. This approach builds upon the function-based approach to treatment but provides a systematic method for identifying candidate PBMs and then quantifying how well those variables predict response to treatment. Unlike traditional behavioral research that focuses on a small number of individuals to demonstrate effects, this approach requires an analysis of multiple cases that includes a sizable number of both responders and nonresponders to a given intervention (ideally, using a sample that is representative of a clinical population). It bears noting that a

research agenda aimed at identifying additional PBMs relies upon knowledge acquired from smaller-scale studies designed to isolate variables related to the mechanisms of problem behavior and putative action treatment, as these may hold the most promise as potential PBMs. In addition to identifying variables that predict response to behavior reduction treatments and determine the conditional effectiveness of treatments, it may be possible to extend these methods to identify PBMs that predict response to skill acquisition programming, and to identify variables that predict response to assessment procedures.

Acknowledgments

Manuscript preparation was supported by Grant R01 HD076653 and from the Eunice K. Shriver National Institute of Child Health and Human Development (NICHD), and U54 HD079123 from the Intellectual and Developmental Disabilities Research Centers (IDRC). The contents are solely the responsibility of the authors and do not necessarily represent the official views of NICHD or IDRC.

REFERENCES

- Asmus JM, Ringdahl JE, Sellers JA, Call NA, Andelman MS, & Wacker DP (2004). Use of a short-term inpatient model to evaluate aberrant behavior: Outcome data summaries from 1996 to 2001. *Journal of Applied Behavior Analysis*, 37, 283–283. 10.1901/jaba.2004.37-283 [PubMed: 15529887]
- Baer DM, Wolf MM, & Risley TR (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91–97. 10.1901/jaba.1968.1-91. [PubMed: 16795165]
- Brown KA, Wacker DP, Derby KM, Peck SM, Richman DM, Sasso GM, ... Harding JW (2000). Evaluating the effects of functional communication training in the presence and absence of establishing operations. *Journal of Applied Behavior Analysis*, 33, 53–71. 10.1901/jaba.2000.33-53. [PubMed: 10738952]
- Carr JE, Severtson JM, & Lepper TL (2009). Noncontingent reinforcement is an empirically supported treatment for problem behavior exhibited by individuals with developmental disabilities. *Research in Developmental Disabilities*, 30, 44–57. 10.1016/j.ridd.2008.03.002. [PubMed: 18467073]
- Carser JE, Quinn JE, Michie CO, O'Brien EJ, McCluggage WG, Maxwell P, ... Gourley C (2011). BRCA1 is both a prognostic and predictive biomarker of response to chemotherapy in sporadic epithelial ovarian cancer. *Gynecologic Oncology*, 123, 492–498. 10.1016/j.ygyno.2011.08.017. [PubMed: 21920589]
- Critchfield TS, & Kollins SH (2001). Temporal discounting: Basic research and the analysis of socially important behavior. *Journal of Applied Behavior Analysis*, 34, 101–122. 10.1901/jaba.2001.34-101. [PubMed: 11317983]
- Davies LE, & Oliver C (2016). Self-injury, aggression and destruction in children with severe intellectual disability: Incidence, persistence and novel, predictive behavioural risk markers. *Research in Developmental Disabilities*, 49, 291–301. 10.1016/j.ridd.2015.12.003 [PubMed: 26765248]
- Didden R, Korzilius H, van Oorsouw W, & Sturmey P (2006). Behavioral treatment of challenging behaviors in individuals with mild mental retardation: Meta-analysis of single-subject research. *American Journal on Mental Retardation*, 111, 290–298. 10.1352/0895-8017(2006)111[290:btocbi]2.0.co;2. [PubMed: 16792430]
- Donaldson JM, Vollmer TR, Krous T, Downs S, & Berard KP (2011). An evaluation of the good behavior game in kindergarten classrooms. *Journal of Applied Behavior Analysis*, 44, 605–609. 10.1901/jaba.2011.44-605. [PubMed: 21941390]
- Drucker E, & Krapfenbauer K (2013). Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalized medicine. *The EPMA Journal*, 4, 7. 10.1186/1878-5085-4-7. [PubMed: 23442211]
- Eldevik S, Hastings RP, Hughes JC, Jahr E, Eikeseth S, & Cross S (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child & Adolescent Psychology*, 38, 439–450. 10.1080/15374410902851739. [PubMed: 19437303]

- Fisher WW, Greer BD, Romani PW, Zangrillo AN, & Owen TM (2016). Comparisons of synthesized and individual reinforcement contingencies during functional analysis. *Journal of Applied Behavior Analysis*, 49, 596–616. 10.1002/jaba.314 [PubMed: 27174563]
- Fritz JN, Iwata BA, Hammond JL, & Bloom SE (2013). Experimental analysis of precursors to severe problem behavior. *Journal of Applied Behavior Analysis*, 46, 101–129. 10.1002/jaba.27 [PubMed: 24114089]
- Ghaemmaghami M, Hanley GP, & Jessel J (2016). Contingencies promote delay tolerance. *Journal of Applied Behavior Analysis*, 49, 548–575. 10.1002/jaba.333 [PubMed: 27449401]
- Greer BD, Fisher WW, Saini V, Owen TM, & Jones JK (2016). Functional communication training during reinforcement schedule thinning: An analysis of 25 applications. *Journal of Applied Behavior Analysis*, 49, 105–121. 10.1002/jaba.265. [PubMed: 26482103]
- Hagopian LP, Fisher WW, Thompson RH, Owen-DeSchryver J, Iwata BA, & Wacker DP (1997). Toward the development of structured criteria for interpretation of functional analysis data. *Journal of Applied Behavior Analysis*, 30, 313–326. 10.1901/jaba.1997.30-313 [PubMed: 9210309]
- Hagopian LP, & Frank-Crawford MA (2017). Classification of self-injurious behaviour across the continuum of relative environmental–biological influence. *Journal of Intellectual Disability Research*. Advance online publication. 10.1111/jir.12430
- Hagopian LP, Rooker GW, & Zarcone JR (2015). Delineating subtypes of self-injurious behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, 48, 523–543. 10.1002/jaba.236. [PubMed: 26223959]
- Hagopian LP, Rooker GW, Zarcone JR, Bonner AC, & Arevalo AR (2017). Further analysis of subtypes of automatically reinforced SIB: A replication and quantitative analysis of published datasets. *Journal of Applied Behavior Analysis*, 50, 48–66. 10.1002/jaba.368. [PubMed: 28032344]
- Hersen M, & Barlow DH (1976). *Single case experiment designs*. New York: Pergamon Press.
- Higginson IJ, & Carr AJ (2001). Using quality of life measures in the clinical setting. *British Medical Journal*, 322(7297), 1297–1300. 10.1136/bmj.322.7297.1297 [PubMed: 11375237]
- Hosmer DW, & Lemeshow S (2000). *Applied logistic regression* (2nd ed). Hoboken, NJ: John Wiley & Sons.
- Hurl K, Wightman J, Haynes SN, & Virues-Ortega J (2016). Does a pre-intervention functional assessment increase intervention effectiveness? A meta-analysis of within-subject interrupted time-series studies. *Clinical Psychology Review*, 47, 71–84. 10.1016/j.jpain.2016.01.473 [PubMed: 27372279]
- Iwata BA, Dorsey MF, Slifer KJ, Bauman KE, & Richman GS (1982). Toward a functional analysis of self-injury. *Analysis and Intervention in Developmental Disabilities*, 2, 3–20. 10.1016/0270-4684(82)90003-9.
- Iwata BA, Pace GM, Cowdery GE, & Miltenberger RG (1994). What makes extinction work: An analysis of procedural form and function. *Journal of Applied Behavior Analysis*, 27, 131–144. 10.1901/jaba.1994.27-131. [PubMed: 8188555]
- Jessel J, Hanley GP, & Ghaemmaghami M (2016). Interview-informed synthesized contingency analyses: Thirty replications and reanalysis. *Journal of Applied Behavior Analysis*, 49, 576–595. 10.1002/jaba.316 [PubMed: 27174653]
- Jessel J, Ingvarsson ET, Metras R, Kirk H, & Whipple R (2018). Achieving socially significant reductions in problem behavior following the interview-informed synthesized contingency analysis: A summary of 25 outpatient applications. *Journal of Applied Behavior Analysis*, 51, 130–157. [PubMed: 29319154]
- Johnston JM, & Pennypacker HS (2009). *Strategies and Tactics of Behavioral Research* (3rd ed.). Hillsdale, NJ: Erlbaum.
- Katsnelson A (2013). Momentum grows to make ‘personalized’ medicine more ‘precise.’ *Nature Medicine*, 19, 249–249. 10.1038/nm0313-249.
- Kern L, Delaney BA, Hilt A, Bailin DE, & Elliot C (2002). An analysis of physical guidance as reinforcement for noncompliance. *Behavior Modification*, 26, 516–536. 10.1177/0145445502026004005. [PubMed: 12205825]

- Kuhn DE, DeLeon IG, Fisher WW, & Wilke AE (1999). Clarifying an ambiguous functional analysis with matched and mismatched extinction procedures. *Journal of Applied Behavior Analysis*, 32, 99–102. 10.1901/jaba.1999.32-99. [PubMed: 10201106]
- Kurtz PF, Boelter EW, Jarmolowicz DP, Chin MD, & Hagopian LP (2011). An analysis of functional communication training as an empirically supported treatment for problem behavior displayed by individuals with intellectual disabilities. *Research in Developmental Disabilities*, 32, 2935–2942. 10.1016/j.ridd.2011.05.009 [PubMed: 21696917]
- Lalli JS, Mace FC, Wohn T, & Livezey K (1995). Identification and modification of a response-class hierarchy. *Journal of Applied Behavior Analysis*, 28, 551–559. 10.1901/jaba.1995.28-551 [PubMed: 16795881]
- La Thangue NB, & Kerr DJ (2011). Predictive biomarkers: A paradigm shift towards personalized cancer medicine. *Nature Reviews Clinical Oncology*, 8, 587–596. 10.1038/nrclinonc.2011.121
- LeBlanc LA, Patel MR, & Carr JE (2000). Recent advances in the assessment of aberrant behavior maintained by automatic reinforcement in individuals with developmental disabilities. *Journal of Behavior Therapy and Experimental Psychiatry*, 31, 137–154. 10.1016/s0005-7916(00)00017-3 [PubMed: 11132117]
- McDonald OG, Li X, Saunders T, Tryggvadottir R, Mentch SJ, Warmoes MO, ... Stauffer KM (2017). Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nature Genetics*, 46, 367–376. 10.1038/ng.3753.
- Miller N, & Neuringer A (2000). Reinforcing variability in adolescents with autism. *Journal of Applied Behavior Analysis*, 33, 151–165. 10.1901/jaba.2000.33-151. [PubMed: 10885524]
- National Cancer Institute (n.d.). <https://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=285747>
- National Research Council. (2011). *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease*. Washington DC: National Academies Press.
- Neef NA, Marckel J, Ferreri SJ, Bicard DF, Endo S, Aman MG, ... Armstrong N (2005). Behavioral assessment of impulsivity: A comparison of children with and without attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis*, 38, 23–37. 10.1901/jaba.2005.146-02 [PubMed: 15898472]
- Okimoto RA, & Bivona TG (2014). Recent advances in personalized lung cancer medicine. *Personalized Medicine*, 11, 309–321. 10.2217/PME.14.19 [PubMed: 25506379]
- Pepe MS (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford, U.K.: Oxford University Press.
- Reed DD, & Kaplan BA (2011). The matching law: A tutorial for practitioners. *Behavior Analysis in Practice*, 4, 15–24. 10.1007/BF03391780 [PubMed: 22649575]
- Richman DM, Wacker DP, Asmus JM, & Casey SD (1998). Functional analysis and extinction of different behavior problems exhibited by the same individual. *Journal of Applied Behavior Analysis*, 31, 475–478. 10.1901/jaba.1998.31-475. [PubMed: 9757585]
- Richman DM, Wacker DP, Asmus JM, Casey SD, & Andelman M (1999). Further analysis of problem behavior in response class hierarchies. *Journal of Applied Behavior Analysis*, 32, 269–283. 10.1901/jaba.1999.32-269. [PubMed: 10513024]
- Ringdahl JE, Vollmer TR, Marcus BA, & Roane HS (1997). An analogue evaluation of environmental enrichment: The role of stimulus preference. *Journal of Applied Behavior Analysis*, 30, 203–216. 10.1901/jaba.1997.30-203.
- Roane HS, Fisher WW, Kelley ME, Mevers JL, & Bouxsein KJ (2013). Using modified visual-inspection criteria to interpret functional analysis outcomes. *Journal of Applied Behavior Analysis*, 46, 130–146. 10.1002/jaba.13. [PubMed: 24114090]
- Scheithauer M, Cariveau T, Call NA, Ormand H, & Clark S (2016). A consecutive case review of token systems used to reduce socially maintained challenging behavior in individuals with intellectual and developmental delays. *International Journal of Developmental Disabilities*, 62, 157–166. 10.1080/20473869.2016.1177925
- Scott SA (2011). Personalizing medicine with clinical pharmacogenetics. *Genetics in Medicine*, 13, 987–995. 10.1097/GIM.0b013e318238b38c. [PubMed: 22095251]
- Skinner BF (1953). *Science and human behavior*. New York, NY: Simon and Schuster.

- Slaton JD, Hanley GP, & Raftery KJ (2017). Interview-informed functional analyses: A comparison of synthesized and isolated components. *Journal of Applied Behavior Analysis*, 50, 252–277. 10.1002/jaba.384 [PubMed: 28304093]
- StataCorp. (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.
- Vaughan ME, & Michael JL (1982). Automatic reinforcement: An important but ignored concept. *Behaviorism*, 10, 217–227. Accessed at: http://www.jstor.org/stable/27759007?seq=1#page_scan_tab_contents
- Vollmer TR (1994). The concept of automatic reinforcement: Implications for behavioral research in developmental disabilities. *Research in Developmental Disabilities*, 15, 187–207. 10.1016/0891-4222(94)90011-6 [PubMed: 7938787]
- Vollmer TR, Borrero JC, Lalli JS, & Daniel D (1999). Evaluating self-control and impulsivity in children with severe behavior disorders. *Journal of Applied Behavior Analysis*, 32, 451–466. 10.1901/jaba.1999.32-451. [PubMed: 10641300]
- Vollmer TR, Iwata BA, Zarcone JR, Smith RG, & Mazaleski JL (1993). The role of attention in the treatment of attention-maintained self-injurious behavior: Noncontingent reinforcement and differential reinforcement of other behavior. *Journal of Applied Behavior Analysis*, 26, 9–21. 10.1901/jaba.1993.26-9 [PubMed: 8473262]
- Watson JB (1913). Psychology as the behaviorist views it. *Psychological Review*, 20, 158–177. 10.1037/h0074428

Figure 1.
 Sample ROC Curve. Line A = perfect classification. Line B = intermediate classification.
 Line C = poor classification. Circled data points identify value of PBioM at which the
 combination of specificity and sensitivity are maximized.

Figure 2.

ROC curve for level of differentiation in the functional analysis and treatment outcome.

Each data point represents the sensitivity and 1 – specificity values obtained for each observed level of differentiation value. The uppermost left data point that is circled, where combined sensitivity and specificity values are maximized, was obtained when the level of differentiation was 63.7% and thus identifies the optimal “cutoff” point for level of differentiation.

Figure 3.

The relation between level of differentiation and treatment effectiveness. Data points above the 80% reduction line (dashed horizontal line) represent responders to treatment. ROC Curve identified optimal cutoff for level of differentiation of SIB in the functional analysis is 63.7% (dashed vertical line). True Positives and True Negatives are data points within shaded areas; False Positives and False Negatives are data points in the unshaded areas.

Table 1

Demographic Information

Variable	Cases Included in Current Analysis									
	All Automatic		Subtype-1		Subtype-2		All Automatic		Subtype-1	
	n	%	n	%	n	%	n	%	n	%
Gender										
Female	24	30.8	14	34.1	10	27.0	8	21.6	6	26.1
Male	53	67.9	27	65.9	26	70.3	29	78.4	17	73.9
Not reported	1	1.3	0	0.0	1	2.7	0	0.0	0	0.0
Age (years)										
3 to 12	33	42.3	16	39.0	17	45.9	16	43.2	8	34.8
13 to 17	17	21.8	8	19.5	9	24.3	8	21.6	6	26.1
>18	28	35.9	17	41.5	11	29.7	13	35.1	9	39.1
Not reported	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Autism										
Yes	30	38.5	16	39.0	14	37.8	17	45.9	14	60.9
No	11	14.1	5	12.2	6	16.2	0	0.0	0	0.0
Not reported	37	47.4	20	48.8	17	45.9	20	54.1	9	39.1
Level of ID										
Mild	5	6.4	4	9.8	1	2.7	2	5.4	2	8.7
Moderate	10	12.8	7	17.1	3	8.1	8	21.6	4	17.4
Severe	13	16.7	10	24.4	3	8.1	8	21.6	7	30.4
Profound	25	32.1	10	24.4	15	40.5	9	24.3	6	26.1
Unspecified	14	17.9	7	17.1	7	18.9	7	18.9	2	8.7
Not reported	11	14.1	3	7.3	8	21.6	3	8.1	2	8.7
SIB topography										
Body-directed	13	16.7	7	17.1	6	16.2	6	16.2	4	17.4
Head-directed	60	76.9	29	70.7	31	83.8	27	73.0	15	65.2
Mouth-directed	1	1.3	1	2.4	0	0.0	1	2.7	1	4.3
Skin-directed	34	43.6	18	43.9	16	43.2	19	51.4	15	65.2

* One individual (Barry, from Ringdahl et al., 1997) had one topography of Subtype-1 ASIB and one topography of Subtype-2 ASIB.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

ROC Identified Cutoff for Level of Differentiation and Treatment Outcomes

ROC Identified Cutoff	Treatment Outcomes			
	All	Responders	Nonresponders	PPV NPV
Differentiation 63.7%	21	18	3	85.7% na
Differentiation <63.7%	16	2	14	n/a 87.5%

Note. Differentiation = proportional difference in average rate of SIB in play condition relative to alone or ignore condition of the functional analysis. ROC = Receiver Operating Characteristic Curve. PPV = Positive Predictive Value, the conditional probability of a positive response to treatment using reinforcement given a level of differentiation at or exceeding the ROC cutoff of 63.7%. NPV = Negative Predictive Value, the conditional probability of a nonpositive response to treatment using reinforcement given a level of differentiation below the ROC cutoff.

Table 3

Subtype Classification and Treatment Outcomes

Group	Treatment Outcomes				PPV	NPV
	All	Responders	Nonresponders			
Subtype-1 ASIB	23	19	4		82.6%	na
Subtype-2 ASIB	14	1	13		na	92.9%
All Datasets of ASIB	37	20	17		na	na

Note. PPV = Positive Predictive Value, the conditional probability of a positive response to treatment using reinforcement given classification as Subtype-1. NPV = Negative Predictive Value, the conditional probability of a nonpositive response to treatment using reinforcement given classification as Subtype-2.