Designs and Analyses of Psychotropic and Behavioral Interventions for the Treatment of Problem Behavior Among People With Intellectual and Developmental Disabilities

Andrea B. Courtemanche, Stephen R. Schroeder, and Jan B. Sheldon
University of Kansas

Abstract
A combination of behavioral and medication-based interventions has been the most effective form of treatment for reducing problem behavior in individuals with intellectual and developmental disabilities. Evaluating the 2 types of interventions in combination and separately may require that researchers adapt methods traditionally used to evaluate drug interventions for individuals without disabilities. Some methodological difficulties that arise when evaluating drug treatments with this population include the withholding of treatment from control groups, identifying large homogeneous samples of participants, predicting individual clinical responsiveness, and many others. The purpose of this article is to summarize the methodological problems that arise when studying drug–behavior interactions among people with intellectual and developmental disabilities and to suggest alternative methods that may ameliorate these issues.

DOI: 10.1352/1944-7558-116.4.315

Treatment of problem behavior is often given the highest priority among people with intellectual and developmental disabilities because of the potential of harm to self or others. A survey in the United Kingdom (Deb, Thomas, & Bright, 2001) reported that 40%–60% of individuals with intellectual and developmental disabilities living in institutional placements display some topography of aberrant behavior, with 11% of those individuals engaging in severe problem behavior, including aggression, property destruction, tantrums, self-injurious behavior, and stereotypies. Surveys in the United States have shown an even broader range of prevalence, depending on the population studied (Baumeister, Todd, & Sevin, 1993; Rojahn & Esbensen, 2002; Rojahn, Schroeder, & Hoch, 2008). Due to the high prevalence of problem behavior in this population, the range of treatments for these difficult behaviors varies; the most common practice, however, is to treat most topographies with psychotropic medications. Since the development of chlorpromazine in 1952, there has been a large increase in the number and variety of psychotropic medications used to treat problem behavior (Lipman, 1970). The number of individuals with intellectual and developmental disabilities who are prescribed medication has received a large amount of attention over the last 40 years, with surveys reporting that from 51% to 57% of individuals in institutions, 26% to 41% of individuals in community group homes, and 22% of individuals in school-based settings are taking at least one form of medication to reduce problem behavior (Aman, Van Bourgondien, Wolford, & Sarphare, 1995; Baumeister et al., 1993; Pyles, Muniz, Cade, & Silva, 1997). The percentage of individuals with intellectual and developmental disabilities taking medications is often related to the individuals’ level of disability, age, gender, other diagnoses,
and residential settings (Rinck, 1998; Schroeder, 1999; Schroeder, Lewis, & Lipton, 1983). The most common classes of medication prescribed by psychiatrists to combat problem behavior include atypical antipsychotics, antidepressants, and mood stabilizers (Aman et al., 1995; Baumeister et al., 1993; Pyles et al., 1997; Unwin & Deb, 2008).

Most often, individuals with intellectual and developmental disabilities are given only medication for the treatment of their problem behavior despite studies as early as 1981 (Sovner & Hurley, 1981) suggesting that a combination of both psychotropic medication and behavioral interventions may be the best option for treating many psychological and behavioral disorders. When experts in psychiatry and psychology were surveyed by Rush and Frances (2000) about their preference for treatment options for reducing problem behavior, most of them reported that their first choice was to try nonmedication based interventions. Psychiatrists have also reported that the number one reason for choosing medication-based treatments was the failure of nondrug (behavioral) interventions. A recent survey in the United Kingdom, however, showed that psychiatrists still rarely report that they have tried supplementing nonmedication-based treatments with medication before abandoning the behavioral treatment (Unwin & Deb, 2008).

Drug–behavior interaction studies assess the effects of a behavioral treatment alone, a pharmacological treatment alone, and the effects of both treatments combined. As previously stated, there appears to be a preference in choosing pharmacological treatments over behavioral treatments when treating aberrant behavior, but Zarcone and colleagues (2008) suggested both could be used simultaneously and in harmony. Evaluating drug–behavior interaction studies with persons with intellectual and developmental disabilities, however, poses a number of different difficulties and concerns. The most immediate consideration involves the choice of research designs: a design that meets the strict methodological requirements of many funding agencies or a design that attempts to directly affect clinical practice. The purpose of this article is twofold: (a) to review the assumptions, advantages, and disadvantages of the different methodologies used to study drug–behavior interactions and (b) to discuss reoccurring methodological and procedural issues that arise when studying drug–behavior interactions with individuals with intellectual and developmental disabilities and to offer suggestions that may ameliorate these issues.

Research Designs: Two Significant Approaches

Group Designs

There are a number of different group designs available to researchers to assess the effects of interventions. Designs may include randomized controlled trials (RCTs) and parallel-dose, factorial, add-on, crossover, and early escape designs (see Evans & Ildstad, 2001, for a comprehensive review). In this article, we focus mainly on RCTs because they are commonly used with the persons with intellectual and developmental disabilities. RCTs are defined as clinical trials that involve at least one test treatment and one control treatment that have been allocated to participants by a random process, with concurrent enrollment and follow-up testing on both the treatment of interest and control treatment (Higgins & Green, 2006). Most RCTs have three consistent measures: premeasures, postmeasures, and follow-up measures. The results of RCTs compare one group with another across the three measures, but they do not analyze the data individually. RCTs estimate the efficacy of the treatments from the percentage of participants who have improved in each group and allow for the calculation of effect sizes to estimate the magnitude of the treatment effects (Chassan, 1976).

Single-Subject Designs

Single-subject designs are another useful method to determine how effective behavioral interventions and psychotropic medications may be at reducing severe problem behavior for an individual, either alone or in combination. Single-subject designs test hypotheses in regards to a single individual, and data are analyzed separately for each participant across different treatment phases using visual inspection (Chassan, 1976). If a single-subject design was used to evaluate different treatment options, each participant would receive a drug and nondrug treatment in a randomized and counterbalanced order. Instead of comparing data across treatment groups as one does when using an RCT, researchers using single-subject designs analyze the level of slope changes between phases by
using continuous recording measures for each participant (Walker, 2008).

---

**Methodological Considerations**

Both of the aforementioned methods have numerous advantages, but both pose multiple methodological considerations when evaluating drug and/or behavioral treatments with people with intellectual and developmental disabilities. Considerations may include the use of control groups, attrition, the heterogeneity of participants and lack of adequate samples, identifying individual treatment options, and cost (Rojahn et al., 2008).

The use of control groups creates an ethical concern of withholding treatment for very severe problem behaviors. Withholding treatment is a problem for most research designs but may become more of a concern when working with dependent populations because of their inability to consent to many of the procedures. The use of single-subject designs or crossover designs (a type of RCT) may alleviate this ethical concern. Single-subject designs allow participants to access all possible interventions in a random and counterbalanced order. Crossover designs randomize participants into groups, with all groups receiving all available treatments assigned in a random sequence. A simple crossover design is an AB–BA design in which participants are randomly assigned to Intervention A or B and then “cross over” to the remaining intervention (Chassan, 1976; Higgins & Green, 2006). In both cases, participants serve as their own control; therefore, the concern regarding the withholding of treatment is virtually eliminated. Unfortunately, because all participants are receiving all treatments, some additional methodological issues may arise. The possibility of treatment-period effects and/or carry-over effects may occur and skew data when participants cross over into the next treatment phase. Wash-out periods may be required when evaluating treatments with long-lasting effects. These confounding variables must be assessed and controlled statistically.

If single-subject designs or crossover designs are used, participants receive all possible treatments, significantly increasing the amount of time these evaluations will take; this may generate high dropout rates. Because single-subject designs do not require a statistical analysis, attrition is not a statistical problem for these designs (Chassan, 1976). Having a small number of participants in a single-subject design, however, may produce less quantifiable results. Analyses such as “intent-to-treat” (ITT) have been developed to address the problem of attrition for RCTs. **ITT analysis** is an approach that analyzes group treatment effects based on the treatment group that was originally assigned rather than the treatment that the group or individual may have actually received. ITT analyses suggest that all participants’ data, complete or incomplete, should be included in the analysis of treatment effects (Mazumdar, Liu, Houck, & Reynolds, 1999). ITT analyses are designed to make RCTs more representative of real-world clinical application because they account for patients or participants who may not have adhered to the treatment regimen or had withdrawn from treatment, which commonly occurs during clinical practice. For example, if an RCT had Groups A and B, an ITT analysis would categorize each participant into one of four groups: (a) participants who were assigned to and completed the Group A protocol, (b) participants who were assigned to but deviated from the Group A protocol or who withdrew from Group A, (c) participants who were assigned to and successfully completed the Group B protocol, and (d) participants who were assigned to but deviated from the Group B protocol or who withdrew from Group B (Wright & Sim, 2003).

An important consideration in the design of studies is the method used to recruit participants and how that method can affect the generalizability of the results obtained. For group designs, recruiting large homogenous samples of people with intellectual and developmental disabilities to participate is extremely difficult. Ideally, RCTs recruit individuals from a homogenous population and then randomly assign participants to a treatment group. The results then can be generalized to other individuals who may fall into the original homogenous sample pool. Unfortunately, because homogenous sample pools of people with intellectual and developmental disabilities are limited and the groups of these individuals who are recruited for RCTs are largely heterogeneous, the generalizability of the results obtained are severely limited (Birmbrauer, Peterson, & Solnick, 1974). In addition, because RCTs only report the mean effect of a treatment for a group, which averages out any individual variability and idiosyncratic responses (Chassan, 1976), the likelihood that the results obtained from the
RCT can be used to predict clinical responsiveness for specific individuals (i.e., ecological fallacy) is further reduced. An internally valid RCT with limited applicability is a lost opportunity to affect clinical practice (Treweek & Zwartenstein, 2009).

On the other hand, single-subject designs may allow researchers to analyze both drug responders versus nonresponders and unique idiosyncratic responses to drug and nondrug treatments (Horner et al., 2005), thus allowing for the development of individualized treatment options for those participating in the research project. It is unclear, however, if the results obtained in single-subject designs can be generalized to other individuals outside of the research study. In addition, it is important to note that single-subject research designs that have failed to identify an effective treatment for those participating are rarely ever published, whereas RCTs often report results for both effective and ineffective treatments.

A final variable to consider is cost. Evaluating behavioral and drug treatments are costly in both time and resources. RCTs are especially costly in that they are usually multisite and require monitoring a large number of participants, consultants, staff members, and other research contributors. It is unlikely that researchers would be able to conduct a large, multisite RCT without some type of external funding. Although single-subject designs may be less costly in regards to funding, they may be just as costly in time. Although single-subject designs recruit a significantly smaller participant pool compared with RCTs, continuous recording and repeated measures for each participant may require a significant amount of time.

**Levels of Evidence: Efficacy Versus Effectiveness**

Prior to choosing a research design, researchers must decide if their goal is to evaluate the efficacy or the effectiveness of the intervention of interest. Efficacy studies evaluate the impact of the intervention under optimal conditions and are likely to be more controlled, whereas effectiveness studies assess the impact of the intervention in real-world settings and may make some adaptations to traditional procedures for use in clinical settings. Efficacy studies are usually done first to determine whether the intervention produces desirable results under optimal conditions; effectiveness studies are subsequently conducted to determine an individual’s responsiveness to the intervention (Tunis, Stryer, & Clancy, 2003).

Clinical drug trials traditionally have four phases of evaluation. Phase I includes evaluating the drug treatment with small numbers of typical participants in a tightly controlled setting. Phase II involves larger numbers of participants from a unique target population, with the drug treatment investigated in a tightly controlled setting using either open or single-blinded procedures. Phase III trials involve studying a drug treatment with a target population under outpatient conditions using double-blinded, placebo-controlled procedures. Phase IV trials may provide information about both the efficacy and effectiveness of certain interventions. Phase IV trials involve open studies investigating the drug’s effectiveness (McCannell & Duff, 1995).

**Practical clinical trials (PCTs)**, a type of Phase IV trial, have been developed to assess the effectiveness of different interventions. PCTs differ from Phase I, II, and, to some extent, III trials in that they assess the real-world applicability of interventions. PCTs evaluate the risk–benefit ratio of the intervention of interest in addition to the cost of using that intervention in clinical practice. PCTs evaluate clinically and socially significant interventions and often compare these interventions with “standard care” or other commonly used interventions rather than comparing the intervention with a placebo or no treatment. In addition, PCTs recruit a more diverse sample of participants from a number of different clinical settings and expand the number of dependent variables to a wide range of different health outcomes (Tunis et al., 2003). The main purpose of PCTs is to evaluate the effectiveness of treatments to have an impact on clinical practice and to help identify treatments for specific individuals.

Due to the number of different methods for evaluating the efficacy and effectiveness of clinical trials, many researchers have attempted to categorize clinical trials into different levels based on the study’s methodological rigor. Nathan and Gorman (2002) suggested classifying all studies into six categories. Type 1 studies would be those that are randomized, controlled trials; include random assignment of participants and blinded assessments; have clear inclusion and exclusion criteria; have an adequate sample size to demonstrate statistical power; and clearly describe their statistical methods. Pelham and Fabiano (2008)
proposed to also include within-subject designs, such as crossover studies, that meet all of the other requirements for Type 1 studies except for the criterion of being an RCT. They argued that excluding well-designed, within-subject studies may result in eliminating more than half of the research investigating behavioral interventions for individuals with intellectual and developmental disabilities. Type 2 studies would be those that lack some of the requirements for Type 1 studies but can provide important information. Type 3 studies would refer to studies that may be methodologically limited (e.g., lacking appropriate blinding procedures), usually use open-trial procedures, and are often used to obtain pilot data. Type 4 and Type 5 studies would involve secondary data analyses (e.g., meta-analyses), and Type 6 studies, including case studies and opinion papers, would be of only marginal value. Despite the numerous ways of categorizing clinical trials, there is no gold standard for assessing the validity of these types of studies (Higgins & Green, 2006). Good judgment is required when assessing the validity of trials. In fact, Higgins and Green (2006) mentioned that none of the levels described above for measuring the validity or quality of trials should be used without caution. There is no direct relationship between the level or type of evidence described above and the degree to which the trial is free from bias. More simply, just because a study may have all of the necessary methodological requirements to classify the study as Type 1, it does not necessarily mean that the independent variable was implemented correctly or with high integrity. In addition, with the introduction of new methodologies like PCTs these levels may need to be reassessed.

**Methodological Criteria**

In 1978, Hollon and Beck reported that only 1 of out over 2,000 studies on hyperactivity was appropriately designed to assess the effects of methylphenidate (a stimulant drug commonly prescribed for individuals with attention-deficit–hyperactivity disorder [ADHD]) with a behavioral treatment compared with either treatment alone. In 1983, Schroeder et al. reported that they were only able to find 1 adequately designed drug–behavior interaction study with people with intellectual and developmental disabilities in peer-reviewed journals. Napolitano et al. (1999) published a review of drug–behavior interaction studies 16 years later and reported that there were only 8 articles, involving 13 participants with intellectual and developmental disabilities, published in peer-reviewed journals that fit the criteria of a true drug–behavior interaction study. In an attempt to clearly define the methodological requirements for conducting clinical trials for participants with and without disabilities, Sprague and Werry (1971) recommended six methodological criteria: (a) placebo controlled, (b) random assignment of participants, (c) double-blind procedures (note: the American Psychological Association prefers the term masked as opposed to blind [American Psychological Association, 2010]), (d) standardized doses, (e) standardized evaluations, and (f) appropriate statistical analyses. Since 1971, several additional criteria have been suggested: functional experimental designs that are either group or single subject (Baumeister et al., 1993; Matson et al., 2000), holding other medications and concurrent behavioral treatments constant (Matson et al., 2000), objective measures of target behaviors with observer reliability (Matson et al., 2000), alternative treatments compared with drug treatment (Sprague & Baxley, 1978), and testing of combined effects of the drugs with alternative treatments (Schroeder et al., 1983). Described below are two additional criteria that we believe are important in conducting these trials, including function-based behavioral treatments and social validity measures.

Methodological criteria serve an important purpose in that they guide researchers in developing their methodological plan, but there are also problems in having too many criteria. We are not suggesting that the criteria mentioned above are all inclusive, that all criteria must be met to be categorized as a well-designed or control study, or that all of the criteria listed are of equal importance. The most commonly cited criteria are those described by Sprague and Werry (1971), and, 40 years later, many researchers and clinicians still only regard randomized, double-blind, placebo-controlled designs with a large number of participants as methodologically sound (for reviews, see Nathan & Gorman, 2003; Rojahn et al., 2008). Since 2010, the U.S. Food and Drug Administration (FDA) has supported the criteria suggested by Sprague and Werry. Although the FDA does not provide precise methodological...
criteria, they suggest that a well-controlled study must have a comparison of participants treated with the drug–behavioral intervention of interest with a suitable control group, free from experimenter and observer expectations (FDA, 2010).

As described previously, we think that it is extremely difficult to meet these criteria when evaluating treatments for individuals with intellectual and developmental disabilities and that researchers will have to compromise on at least one of the these criteria. In the following sections, we would like to address each of the criteria mentioned above and make suggestions that may help researchers when developing research protocols that involve evaluating drug and behavioral treatments for individuals with intellectual and developmental disabilities.

**Placebo Controlled**

A placebo is a totally inert substance used as a control condition to compare with the treatment condition (Chassan, 1976). Interestingly, Sandler (2005) reported that placebo effects could occur in individuals with intellectual and developmental disabilities; therefore, it is important for researchers to fully consider the issues involved when using a placebo. As mentioned previously, placebo conditions may pose ethical concerns if they require withholding treatment from those in need. March et al. (2004) suggested that a placebo should only be used if it is fully justified. In other words, placebo conditions should only be used if there is no substantive risk to the participants. A placebo condition may be used if the treatment being studied is more high risk than using a placebo, to help identify placebo responders, and only if the placebo control will be used for a short amount of time (March et al., 2004). An additional point raised by Sandler (2005) is that participants must be adequately informed about the possibility of being assigned to a placebo group and freely consent to participate. This is often difficult to achieve when working with participants who cannot consent or assent to procedures. Scahill et al. (2009) have also suggested that a placebo control may be avoided if the efficacy of the drug treatment has already been demonstrated in previous studies. Another alternative would be to use placebos as the first treatment phase in single-subject designs, preceding the first and lowest dose of the drug being studied (Thompson, Hackenberg, & Schaal, 1991). Although this method only requires that participants be exposed to a placebo for a short amount of time, using this method also introduces another possible confound in that it introduces a single-blind effect (discussed below). Some additional alternatives to using placebo-control conditions may be to use the best proven existing intervention, to use the best comparison treatment or nontreatment available, or to attempt to neutralize treatment expectations (e.g., using blinding procedures; Normile, 2008; O’Leary & Borkovec, 1978). Note that each adaptation from the true placebo control may create additional complications or confounds for investigators. Investigators must weigh the benefits of the adaptations described above, but it seems possible to conduct a well-controlled study without using a standard placebo control, especially when working with dependent populations.

**Random Assignment**

Random assignment involves randomly assigning samples of individuals who are part of the target population to different treatment options (Thompson et al., 1991). Random assignment allows researchers to evaluate if the treatment option or some other factors may have caused the treatment outcome (Walker, 2008). It is important to minimize the differences between the individuals in each group to limit bias in the construction of treatment groups (Chassan, 1976). True randomization involves treatment assignment based on a random process such as random numbers tables. Treatment assignment based on coin flips, participant social security numbers, medical record numbers, or other quasi-random processes are not true randomizations and are often considered controlled clinical trials rather than RCTs (Higgins & Green, 2006). Controlled clinical investigations should not use the same random sequences of treatments repeatedly for multiple studies because they lose randomness (Chassan, 1976). Although participants are not randomized into groups with single-subject designs, randomizing and counterbalancing the order of treatments can allow assessment of carry-over, placebo, and order effects.

**Double Blind**

A double-blind investigation is one where the researchers attempt to reduce both patient and experimenter expectations, by “blinding” both
the patient and the experimenter to the order of treatments (Chassan, 1976). Single-blind procedures only blind the participants to the order of treatments, whereas open trials do not blind either party. Double-blind procedures are considered the gold standard. Without a double-blind study, the data are automatically biased. Walkup, Labellarte, and Riddle (1998) suggested that published, unblinded clinical trials create the impression that a drug treatment was effective and that this is a serious limitation of the reported results. At least one research study, however, has shown that having unblinded treatments or a bias toward one type of treatment may not affect the outcome of the study (Cole, 1962). Cole, for example, reported that although researchers favored pharmacologic therapies over electroconvulsive therapy (ECT) during a nonblinded comparison, ECT was reported to be the more effective therapy. In addition, researchers have suggested that double-blind conditions can easily be broken (Sprague & Werry, 1971), and the side effects that are associated with drugs make it extremely difficult to maintain double-blinded procedures (Barlow, Nock, & Hersen, 2008).

Many individuals participating in double-blinded studies are capable of guessing which treatment they or others are receiving. Researchers have shown that physicians are better than chance at guessing when participants are receiving a drug treatment (Rickels, Lipman, Fisher, Park, & Uhlenhuth, 1970) and that psychiatrists were 90% accurate at guessing when participants were receiving a drug treatment (Sprague & Werry, 1971). McAdam and colleagues (2002) reported that community members without professional or personal experience with drug research or individuals with intellectual and developmental disabilities were able to discriminate differences in participants’ behavior when the participants were receiving a drug or placebo treatment.

Double-blind procedures can be very difficult to use with people with intellectual and developmental disabilities and the types of treatments evaluated. As asserted by Cole (1962), it may be difficult to blind treatments if treatment options are distinctly different from each other. For example, there is a noticeable, distinct difference between a drug treatment and a behavioral treatment; thus, keeping participants and other researchers blind to current treatments can be difficult. In addition, when working with persons with intellectual and developmental disabilities, many caregivers and professionals are involved in their daily care and treatment. Keeping all of these individuals blind to treatments can pose additional difficulties. One way to account for staff and investigators becoming unblinded is to ask whether they became unblinded at the end of each participant’s participation. Additional analyses could be conducted with participants whose observers became unblinded compared with those participants whose observers remained blind to treatments throughout the entire study.

Single-blind procedures are an alternative to double-blind procedures when working with the persons with intellectual and developmental disabilities. Single-blind procedures keep the spirit of double-blind procedures by keeping participants and observers unaware of treatment conditions but allow the experimenter to monitor data throughout the course of the clinical investigation. It is nearly impossible to program set drug-phase lengths (e.g., 4 weeks) prior to starting the study, and single-blind procedures offer the option of allowing the researchers to change phases based on the participants’ performance, not a preset phase length that may be inappropriate for some participants. Chassan (1976) reported single-blind procedures were used more often with participants who were treated on a long-term basis. In addition, single-blind procedures used in drug clinical trials are often paired with direct behavioral observation with observer reliability, which have been shown to be free from experimenter bias (Singh & Beale, 1986; Towns, Singh, & Beale, 1984).

Open trials may also provide researchers with some important information. Open trials may be useful for collecting pilot data, may show idiosyncratic or unique adverse effects of drugs, may provide information about the drug’s time course, and may provide important information about the drug’s pharmacokinetics. An open trial may be useful when identifying a clinically effective dose for a small pool of participants. Valuable information can be obtained from conducting open-trial research; thus, accumulation of clinical data should not be limited to controlled clinical trials (Chassan, 1976).

**Standardized Doses**

Sprague and Werry (1971) suggested using a standardized dose (mg/kg) to obtain information about the effects of a drug treatment. There are several ways of achieving this. Some basic designs
allow investigators to choose a fixed individualized dose for each participant or choose a fixed dose that may be effective for the average participant. Because fixed doses may affect each individual differently, Scahill et al. (2009) suggested an additional standardized method of dose adjustment where the researcher or clinician can increase or decrease the dose for each participant to manage any adverse side effects. Although the dose should be somewhat standardized, several different doses should be evaluated. Different doses of the same drug may affect different classes of behavior (Sprague & Sleator, 1977). Schroeder et al. (1983) suggested that the best way to work with standardized doses is to use several different drug doses with different doses of behavioral intervention used in a crossover design in which each participant receives all doses of the drug and behavioral treatment in a randomized and counterbalanced order. For example, Fabiano and colleagues (2007) evaluated a placebo, three doses of methylphenidate (0.15 mg/kg, 0.30 mg/kg, and 0.60 mg/kg), and three doses of the same behavioral intervention for the treatment of children with ADHD using a crossover design during a summer treatment program. The doses of behavioral intervention were no behavioral intervention; a low-intensity intervention that included the use of a point system, activity rules, and social and sports skills; and a high-intensity intervention that included all of the components of the low-intensity intervention plus a time-out procedure and social reinforcement. This study evaluated each dose of methylphenidate with each dose of behavioral intervention to identify the most effective combination for each participant.

**Standardized Evaluations**

Standard evaluations should be taken on all dependent measures. Sprague and Werry (1971) suggested using measurement devices that have been used previously, have some empirical basis, and are reliable and valid. In addition to these characteristics, other researchers have suggested that measurement systems be relevant, practical, sensitive, safe to use, and ethical (Werry, 1978; Williams & Saunders, 1997). We advocate the addition of direct-observation procedures to identify changes in participants’ behavior. Direct-observation procedures require that investigators measure behavior by directly observing behavior rather than relying on someone’s (e.g., staff’s) recollection or opinion of whether the behavior occurred and the characteristics (e.g., frequency, duration, severity) related to the behavior. To identify changes in behavior that are attributed to drugs, Zarcone and colleagues (2008) suggested that when doing direct observations, it is important that operational definitions be used. Operational definitions and direct observations allow at least two independent observers to simultaneously and independently record dependent variables across sessions and participants to assess reliability (Horner et al., 2005). In addition, one should assess the validity of the definitions of the behaviors being observed as well as ensure that the direct observations are scheduled so that representative samples (i.e., across times and contexts) of the participants’ behavior are captured during each phase of treatment.

It is often difficult to conduct direct observations of participant behavior because of the number of staff needed, time, effort, and expertise required to conduct several observations across many different participants. To alleviate this problem, rating scales and medical charts are often used to monitor changes in behavior. Rating scales can provide some important information about changes in the intensity of behavior and can provide information about the variety of settings in which the behavior occurs. The drug effects recorded on medical charts or rating scales, however, may not provide objective and quantifiable information. Rating scales should only be used if they have been standardized and they have been shown to be sensitive to medication changes (Napolitano et al., 1999). If possible, rating scales should be supplemented with direct observations. Direct observations may provide investigators with information that rating scales may not, including specific information about the environment and context in which the behavior occurs and may be more sensitive to small changes in the frequency of behaviors. Rating scales may not be able to capture small changes in the frequency of behavior, but may provide quantifiable information about changes in the intensity or magnitude of behavior. Using both rating scales and direct observations may allow investigators to document and quantify any possible drug effect (Napolitano et al., 1999; Zarcone, Napolitano, & Valdovinos, 2008).
Drug and Behavioral Treatments Evaluated

When studying drug–behavior interactions, four types of evidence are needed: (a) evidence of each treatment’s effectiveness alone, (b) an understanding of how each treatment works alone, (c) evidence of the treatments’ combined effects, and (d) an understanding of how two interventions work together (Klerman, 1975). Understanding the interaction between a behavioral intervention and a drug intervention requires an estimation of the effects of both treatments separately as well as in combination. If the effect of the interventions combined equals the sum of the individual effects of each treatment, it is called addition; if the combined effect is greater than the sum of the individual effects, it is called potentiation; and if the combined effects are less than the individual effects it is called inhibition (Schaal et al., 1983; Uhlenhuth, Lipman, & Covi, 1969; Williams & Saunders, 1997). The individual or combined treatments should be presented in a random order and should be counterbalanced across participants. In addition, data collection during each treatment phase should be long enough to allow for any carryover effects from prior treatments to diminish so that trend effects can be distinguished from treatment effects (Chassan, 1976).

Function-Based Behavioral Treatments

Drug treatments may affect behavior by changing the environmental variables that control behavior (Branch, 1984). Behavior may be reinforced or punished by a number of different environmental contingencies. Drugs may change the reinforcing or punishing values of these variables. In addition, between-subjects variability of drug effectiveness may suggest that different environmental and neurological differences can affect behavior (Schaal & Hackenberg, 1994). Due to the large role that environmental variables have in controlling behavior, it may be necessary to assess the effect those variables have during treatment phases. Functional assessment procedures attempt to identify the environmental variables that control behavior. Functional assessment procedures may help predict potential positive responders and are essential to developing function-based behavioral treatments.

Functional assessment methodology involves observing individuals in their natural environment, interviewing caregivers, and manipulating environmental contingencies to develop treatments for problem behavior. The standard analog functional analysis procedure (manipulation of environmental contingencies) was developed by Iwata et al. (1982) to test experimentally, in an analog environment, the functions of different environmental contingencies (i.e., access to attention, access to preferred items, self-stimulation, or to escape demands) in a brief, counterbalanced design. The literature now includes over 450 studies (Kahng, Iwata, & Lewin, 2002) and has shown functional analysis to be an effective strategy for designing behavioral interventions over the past 30 years. In fact, functional behavioral assessments (FBAs) are now required by federal law in all school-based behavioral interventions for children with intellectual and developmental disabilities (Individuals with Disabilities Education Act, 1997).

Schaal and Hackenberg (1994) suggested using functional assessment methodology when evaluating pharmacological treatments to treat problem behavior. Researchers have demonstrated that drug treatments can directly affect and change environmental consequences that maintain problem behavior. Zarcone et al. (2001), for example, reported that risperidone, an atypical antipsychotic, not only reduced the occurrence of destructive behavior but affected the behavioral function of destructive behavior in some participants. For example, 1 participant’s functional analysis results suggested destructive behavior was maintained by access to tangible items and to escape from demands during baseline and placebo phases. After the introduction of risperidone, the participant’s functional analysis results suggested that the behavior was now only maintained by access to tangible items. Valdovinos et al. (2009) reported similar results when they used functional assessment methodology to assess the effects that different combinations of stimulant and antipsychotic medications had on the behavioral functions of participants’ problem behavior. All participants’ problem behavior functions changed when different doses and different combinations of medications were evaluated. Differential effects in the outcome of drug treatments may be an indication that problem behavior of similar topographical forms in different individuals may be maintained by different consequences (Schaal & Hackenberg, 1994). Functional assessment technology is crucial to identify those maintaining reinforcers.
It is important to note that functional assessments and analyses are resource intensive. We are not suggesting that a traditional, intensive assessment (interviews, direct observations, and manipulation of environmental variables; Iwata et al., 1982) be conducted during all phases or even with small medications changes. Rather, we believe that most professionals and caregivers who work with individuals with intellectual and developmental disabilities are using some type of modified FBA as part of the individuals’ treatment packages (e.g., antecedent–behavior–consequence data, ecobehavioral assessment) and that these FBAs can be used to assess behavioral function after medication changes. We believe that functional assessments (traditional or modified) should only be conducted by professionals who have received the appropriate behavioral instruction to conduct these assessments. Thus, it is not the responsibility of the psychiatrist or pharmacologist to conduct these analyses. We suggest that there be a multidisciplinary approach to treating problem behavior (i.e., biobehavioral approaches) with psychiatrists, pharmacologists, and behaviorists working jointly. Most individuals need multiple interventions (i.e., drug and behavioral) to reduce problem behavior, especially if that problem behavior is controlled by both biological and environmental events. Although functional assessments are costly, we believe that conducting functional assessments will increase the likelihood that supplemental behavioral intervention will be effective by assessing how the psychotropic medication may have altered the behavioral function of the problem behavior.

Social Validity

Social validity measures report whether the consumers of the procedures find the procedures to be acceptable and effective (Horner et al., 2005; Wolf, 1978). These measures may predict whether the consumers will continue to use the intervention for the treatment of people with intellectual and developmental disabilities and may predict whether consumers are satisfied with the long-term effects of the interventions (Poling & LeSage, 1995). Most important, social validity measures indicate whether the study is serving the needs of those participating or if the study is simply meeting the criteria of a well-designed study (McAdam et al., 2002). Social validity measures are altered slightly when treatments are evaluated for individuals with intellectual and developmental disabilities. Because the consumers themselves often cannot communicate whether they thought the treatments were acceptable and they will continue to use them, consumer caregivers are surveyed instead. Caregivers may be asked questions regarding whether they were satisfied with the outcomes of the study, whether they thought the dependent measures and assessments were appropriate, and whether they would recommend the procedures or participation in the study to another parent or caregiver (Aman & Wolford, 1995; Tierney et al., 2007). In addition, it may be helpful to provide caregivers with open-ended questions about the research project rather than close-ended questions (e.g., Likert-type scales) to promote caregiver comments and suggestions (Aman & Wolford, 1995).

Unfortunately, there is no standard methodology for assessing social validity with individuals with intellectual and developmental disabilities who lack communicative abilities. More recently, researchers have assessed consumers’ preference for specific interventions by incorporating choice into the research methodology. Participants may be asked to choose between two different behavioral interventions, each represented by a different colored mat or room (see Hanley, Piazza, Fisher, Maglieri, 2005, for an example). The intervention that the participants choose most frequently is considered the most preferred. Although this can be used easily with behavioral interventions, assessing choice presents challenges when evaluating preference for drug treatments.

Conclusions

Investigating the effects of behavioral and pharmacological treatments alone and in combination may be extremely beneficial in treating severe problem behavior for individuals with intellectual and developmental disabilities. A significant amount of research has been conducted evaluating one treatment option alone, but true drug–behavior interaction studies are rarely seen in peer-reviewed literature. As described above, there are many different methodological procedures available to investigate the effects that psychotropic drugs and behavioral interventions have on problem behavior. There is, however, no one set of procedures that is universally accepted. Although the 6 methodological criteria discussed by Sprague and Werry (1971) are the
most commonly referenced, we have discussed the difficulties and constraints when using these criteria with people with intellectual and developmental disabilities. In addition, it is inappropriate to consider a group, placebo-controlled, double-blinded design as the only methodologically sound research design because it eliminates many important research studies that may make significant contributions to the field of intellectual and developmental disabilities. In our opinion, researchers should consider many issues when conducting this type of research and methodological options should be expanded based on the research question and target population. It is possible to make adaptations to Sprague and Werry’s (1971) original methodological criteria and still have both a well-designed and clinically significant study.

Although these concerns have been prevalent for at least 2 decades, administration and funding agencies have been slow to adopt the changes discussed. According to Treweek and Zwarenstein (2009), the FDA still only associates RCTs with study excellence, thus favoring trials that lack the information needed to affect health care decisions. In addition, many clinical trials that are funded are designed to meet FDA standards. More than 90% of the clinical trials funded would be classified as either a Phase I, II, or III trial, with the remaining 10% being Phase IV trials. Many of the adaptations and studies discussed in this article, especially PCTs, would be classified as Phase IV trials (Getz & Zisson, 2003). The National Institutes of Health (NIH) has begun to fund some PCTs (Tunis et al., 2003), but no PCT-like studies have reportedly been funded evaluating drug and behavioral treatments for problem behavior in persons with intellectual and developmental disabilities. In 1983, Schroeder et al. reported that, in the United States, less than 1% of federal funds for psychotropic drug studies at the National Institute of Mental Health were allocated to research on intellectual disabilities. In 2009, the NIH allocated 2% of funds toward research on intellectual and developmental disabilities (NIH, 2010). Interestingly, people with intellectual and developmental disabilities have been called the most medicated segment of our society. When it comes to the use of psychotropic drugs, perhaps more funds should be dedicated to finding appropriate and effective treatments for this dependent population.

References


© American Association on Intellectual and Developmental Disabilities


Received 2/15/2010, accepted 12/23/2010.

Editor-in-Charge: Frank Symons

Andrea B. Courtemanche, University of Kansas, Applied Behavioral Science, Lawrence, KS 66045.

E-mail: acourtem@ku.edu