Technical Report to the National Coalition for Access to Autism Services (NCAAS):
Department of Defense Reports to the Congress Regarding the Autism Care demonstration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Ira L. Cohen, Ph.D.

Key Findings

In this technical paper, Dr. Ira Cohen analyzes a multiyear Department of Defense (DoD) evaluation to measure the progress of military children with autism and finds the DoD's conclusions to be substantively flawed and based on the incompetent interpretation and analysis of the collected data. He writes that "the conclusions one draws from an analysis are only as good as the quality of information" and that multiple issues with how the DoD scored and analyzed the data "render the conclusions drawn by DoD uninterpretable".

In the paper, Dr. Cohen examines data collected by the DoD from the PDD Behavior Inventory (PDDBI) and the DoD's associated analysis. The PDDBI was created by Dr. Cohen and is an assessment tool used by practitioners worldwide to evaluate children with autism as well as their treatments including Applied Behavior Analysis (ABA) and other interventions. When properly used, this autism assessment tool is trusted by healthcare professionals all over the world. PDD, which stands for Pervasive Developmental Disorder, is another term for Autism Spectrum Disorder (ASD).

ABA is a therapy based on the science of how people learn and aims to decrease behaviors that adversely impact learning, or are harmful, and increase those behaviors that are helpful and positive. ABA has been successfully used to treat children with autism for decades and can help improve attention, academics, and social skills along with improving language and basic communications skills. ABA is considered an *"evidence based"* best practice treatment by the U.S. Surgeon General, and the American Academy of Pediatrics has noted that "most evidence-based treatment models are based on the principles of ABA."

Dr. Ira Cohen analyzes six reports from the DoD to Congress from FY 2018 to 2020[*] that measure the treatment progress of military children with autism under a program run by TRICARE, the DoD's health insurance for military families. The DoD is required to report both quarterly and annually to Congress on the DoD's ABA treatment program.

"The purpose of this paper is to evaluate how the DoD has used the PDDBI to determine if the children receiving ABA services improved over time," wrote Dr. Cohen. The reports "utilize the PDDBI for making determinations about the efficacy of TRICARE's Autism Demonstration and... upon which the DoD draws its conclusions."

An example of Dr. Cohen's conclusion that the DoD has misused the PDDBI assessment tool is *The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress Fourth Quarter, Fiscal Year 2018* in which the DoD discarded 90% of the 14,700 patient assessments in their original test sample saying that there was missing information and subsequently drew conclusions from only 1,577 children.

On this flawed data, Dr. Cohen wrote that the **"most glaring is the total lack of understanding as to how the PDDBI is scored and interpreted and the obvious failure to read the manual**. This led to the elimination of vast numbers of cases, many of whom may have been responders, because of the incorrect assumption that a score of "0" means that the item was 'missing'. Instead, it means that the behavior was not seen."

Dr. Cohen found similarly flawed data in all six reports and that the DoD's conclusions that military children with autism were not making progress with Applied Behavior Analysis "cannot be justified due to profoundly serious limitations" with how the PDDBI was scored and how the resulting data were analyzed.

Dr. Cohen is one of the world's foremost scientific authorities and is the author/co-author of more than 100 peer-reviewed scientific and medical papers on autism.[1]

1 * Report to Congress: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress Fourth Quarter, Fiscal Year 2018

Report to the Committees on Armed Services: The Department of Defense Comprehensive Autism Care Demonstration Annual Report June 2019

Report to Congress: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress Second Quarter, Fiscal Year 2019

Report to Congress: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to the Armed Services Committees, Third Quarter, Fiscal Year 2019

Report to Armed Services Committees: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress, Fiscal Year 2019

Report to the Committees on Armed Services of the Senate and House of Representatives: The Department of Defense Comprehensive Autism Care Demonstration Annual Report 2020

Background

Ira L. Cohen, PhD, specializes in autism, its diagnosis and behavioral assessment. He is the principal author of the PDD Behavior Inventory, the analytical tool that the Department of Defense used to draw conclusions about the effectiveness of applied behavior analysis (ABA) in six Department of Defense Reports to Congress on the Comprehensive Autism Care Demonstration.

Dr. Cohen received his PhD in psychobiology from Rutgers University in 1974 under the supervision of Dr. Michael R. D'Amato. He was an NIH post-doctoral fellow in experimental psychopathology and behavior therapy at NYU Medical Center from 1975-1977. He was then employed as a research psychologist at NYU Medical Center/Bellevue Hospital under the supervision of Dr. Magda Campbell, where he worked with children with autism from 1977-1980. There, he developed objective rating scales, a carpet-based activity monitor, and operant conditioning procedures for quantifying behaviors in the children. He joined the Institute for Basic Research in Developmental Disabilities in 1981 as part of the Autism Demonstration Project under the supervision of Dr. Erwin Friedman. There, he served as a research scientist and, later, as head of the Division of Behavioral Assessment and Research and chair of the Department of Psychology. He is currently Visiting Research Scientist at the Institute. He completed his Autism Diagnostic Interview training and certification in June 1999 under Christopher Smith of Mt. Sinai Hospital. He then completed his Autism Diagnostic Observation Schedule training and certification in August 1999 under Catherine Lord at the University of Chicago. He is a member of the International Society for Autism Research (INSAR) and a member of the editorial board of the Journal of Autism and Developmental Disorders.

Dr. Cohen is the author and coauthor of over 100 peer-reviewed publications on autism. As stated above, Dr. Cohen is the coauthor of the PDD Behavior Inventory (PDDBI), and the author of the PDD Behavior Inventory Screening Version and the Autism Spectrum Disorder Decision Tree (ASD-DT). Dr. Cohen's educational credentials are as follows: BA in psychology, Temple University; MS in psychobiology, Rutgers University; PhD in psychobiology, Rutgers University; NIH post-doctoral fellow, Experimental Psychopathology and Behavior Therapy, NYU Medical Center.

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Introduction

The Department of Defense (DoD) provides health-related services to their active duty (ADFM) and non-active duty (NADFM) family members through TRICARE. One of these services, Applied Behavior Analysis (ABA), is provided for children with a diagnosis of Autism Spectrum Disorder (ASD), which I briefly describe below as it will play a role in my later comments.

What is ASD?

ASD is a behaviorally and etiologically (i.e., causally) heterogeneous neurodevelopmental disorder characterized by abnormalities in the development of reciprocal social interaction and communication and by repetitive and/or ritualistic behaviors. It differs from intellectual disability in that the problems with socialization and communication in autism are qualitatively, versus quantitatively, different from typical development. That is, children with intellectual disability are delayed in the development of social-communication skills, but children with ASD display behaviors not usually seen in children with intellectual disability or in young, typically developing children. For example, affected individuals often do not establish eye contact appropriately when others are talking to them or often do not show interest in sharing attention with others. As young children, they usually do not use gestures to communicate but, instead, guide others to what they want. Vocally, they may be mute or may echo meaninglessly what others say to them, insist on talking to others about topics of conversation that are of little interest to others, and communicate with an unusual vocal intonation style. Repetitive or ritualistic behaviors often displayed include, but are not limited to, hand flapping or bizarre finger movements, unusual walking patterns, meaningless spinning of self or objects, excessive fascination with parts of an object, or severe self-injury (Cohen, 2006). Many of these behaviors change with age, with the severity of such behaviors varying across individuals, influenced, in part, by genetic (Cohen et al., 2011; Cohen, et al., 1991a, b; Cohen., et al, 2016), and health-related problems (Restrepo, et al., 2020).

Intervention for these children can be daunting since, as described, there are multiple issues that need to be addressed including learning to establish eye contact, paying attention to the instructor, learning how to play with toys (in younger children), learning how to socially interact with and share interests with others, learning to generalize skills to various situations (Cohen, 2006), and, importantly, learning what words and sentences mean and learning to use those words to express their needs and desires to others in meaningful and adaptive ways. Such intervention is time-consuming and usually involves parents as co-therapists to facilitate skill acquisition. One of the most common interventions used for these children is applied behavior analysis (ABA).

ABA is an outgrowth of basic laboratory research on how behaviors are acquired, how they are maintained over time and how they can be modified, typically using single-subject research designs that recognize the importance of individual differences, establishing causality and replicating findings across subjects/cases (Sidman, 1960). ABA methods have been used in clinical, industrial, environmental, and educational settings. While there exist many different forms of ABA, for children

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

with autism, common characteristics include careful documentation of the child's skills and problem behaviors expressed in observable terms such that the behaviors can be reliably measured (e.g., percentage correct responses or the frequency of occurrence of self-destructive behaviors). Based on these observations, goal plans are derived, and procedures are implemented to help the child acquire skills and to replace problem behaviors with behaviors that serve the same function but are more adaptive.

The types of behaviors to be taught can be broken down into two areas:

1) Behaviors that will help to mitigate the "core" deficits of ASD: lack of social skills, poor language comprehension/use of language for communication with others, and repetitive behaviors/restricted interests. Here, therapists may focus on teaching the child to imitate, to establish eye contact, to respond to his/her name, to point to communicate needs or to express interest, to understand and use words correctly, to role-play, and to learn to take turns; and

2) Behaviors that may be problematic in ASD but are not unique to it such as poor self-care skills (e.g., problems with tooth brushing, dressing self, chewing and swallowing, toileting, etc.). These also include behaviors that are "challenging" to manage such as aggressiveness, self-injury, anxiety, and hyperactivity, often addressed by teaching the child alternative strategies to communicate his/her needs based on the results of "functional analyses" of the motivation for these challenging behaviors.

These interventions are often derived from published research and can be demonstrated in any given child by appropriate use of single-subject designs. If successful, the interventions are taught to caregivers to enhance the child's learning and generalization of skills. Parents are also taught how to monitor their child's progress toward achieving those goals.

DoD is required to document to congress, quarterly, the effectiveness of the ACD in children receiving such ABA services (Report to Congress: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress Fourth Quarter, Fiscal Year 2018). One of these outcome measures is the PDD Behavior Inventory or PDDBI, developed by the author (Cohen and Sudhalter, 2005).

The purpose of this paper is to evaluate how the DoD has used the PDDBI to determine if the children receiving ABA services improved over time. To this end, I will focus on the following reports, all of which utilize the PDDBI for making determinations about the efficacy of TRICARE's Autism Demonstration and which it appears is the analytical basis of the program's effectiveness and upon which the DoD draws its conclusions:

I.   Report to Congress: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress Fourth Quarter, Fiscal Year 2018;
II. Report to the Committees on Armed Services: The Department of Defense Comprehensive Autism Care Demonstration Annual Report June 2019;
III. Report to Congress: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress Second Quarter, Fiscal Year 2019;
IV. Report to Congress: The Department of Defense Comprehensive Autism Care Demonstra-

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

tion Quarterly Report to the Armed Services Committees, Third Quarter, Fiscal Year 2019;

V. Report to Armed Services Committees: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress, Fiscal Year 2019; and

VI. Report to the Committees on Armed Services of the Senate and House of Representatives: The Department of Defense Comprehensive Autism Care Demonstration Annual Report 2020

Before commenting on the above reports, I will first describe what the PDDBI is and how it is to be used.

What is the PDD Behavior Inventory (PDDBI)?

The PDDBI (Cohen et al., 2005) is a rating instrument designed to be completed by informants who know the child well, usually parents and teachers. At the time the PDDBI was developed, the umbrella term for Autism was "Pervasive Developmental Disorder", hence the PDD abbreviation. The PDDBI comes in two forms, one for parents and one for teachers/other professionals. It was constructed, *a priori*, in a hierarchical manner, and divided into two behavioral *dimensions*: (a) Approach Withdrawal Problems (AWP) assessing atypical core behaviors seen in Autism Spectrum Disorder (ASD) as well as other behavioral concerns seen in other populations (e.g., fears, hyperactivity, sleeping problems, self-injury, aggression); and (b) Receptive/Expressive Social Communication Abilities (REXSCA) assessing skills lacking in ASD such as non-verbal social skills, and receptive and expressive language abilities. Each behavioral dimension is composed of behavioral domains best reflecting that dimension. Brief descriptions of the domains and other scores are presented in Table 1 from Cohen et al. (2016).

Abbreviation	Description and characteristics
AWP	Approach-withdrawal prolems dimension. Higher domain T-scores indicate greater severity
SENSORY	Sensory/percepual approach behaviors-staring at objects, pica, repetitive toy play, hand flapping, etc
RITUAL	Ritualisms/resistance to change - carrying out rituals or indicating dissatisfaction with a change in the enviroment or routine
SOCPP	Social pragmatic problems-problems reacting to the approaches of others, understanding social conventions, or initiating social interactions
SEMPP	Semantic/pragmatic problems-echolalia, perseverative language, unusual voice quality, etc
AROUSE	Arousal regulation problems - emotional constriction, hyperactivity, sleeping problems, etc.
FEARS	Specific fears - fears and anxieties associated with withdrawal from social or asocial stimuli
AGG	Agressiveness - Agressiveness toward self or others and assoaciated negative mood states
REXSCA[a]	Receptive/expressive social communications abilities dimension. Higher domain scores indicate increasing levels of competence
SOCAPP	Social approach behaviors - non-vocal social communication skills such as paying attention, joint attention, effective use of gesture, imaginative skills, social play skills, imitation skills, etc.
EXPRESS	Expressive language - ability to speak sounds associated with the English language as well as competence with grammar, tone of voice, and conversational pragmatics
LMRL	Learning, memory, and receptive language - memory for locations and movement sequences, understanding possesives, prepositions, adverbs, etc.
AUTISTM/C	Autism composite - a measure of lack or appropiate social communication skills along with repetitive / ritualistic behaviors

[a] Each of these domains in the REXSCA dimension is highly correlated with tested IQ [Pearson r (n=76) ranging from 0.63 to 0.77] and with the Vineland Communication Domain score [Pearson r (n=238) ranging from 0.52 to 0.69] (Cohen and Sudhalter 2005)]

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Each domain is comprised of various items describing behaviors scored on a "0" to "3" scale where "0" means the behavior is not displayed and "3" means the child usually or typically shows the behavior. There are 188 items in the parent version and 180 items in the teacher version of the PDDBI. An example is as follows:

> "Item 20: Repetitively moves objects back and forth (e.g., turns pages back and forth, opens and closes doors, turns lights on and off)". For this item, a score of "0" means the informant does not see the child engaging in this repetitive behavior while a "3" means that the informant sees the child doing this quite often.

For domains describing language problems or language skills, persons scoring those domains are instructed to score all the items as "0" if the child cannot speak (Cohen and Sudhalter, 2005, p. 21; more on this below).

The original PDDBI covers ages 1.5-12.5 years and, for each domain, generates T -scores (i.e., the average score for children with ASD is set at 50) that take the age of the child into account. These are generated by summing the scores for the various domains (known as the "raw scores") and then converting these to T-scores by consulting the relevant tables in the PDDBI Manual or by using on-line scoring. The T -scores are normally distributed (see below). By definition, scores less than or greater than 50 indicate greater deviations from a typical case of autism.

The original "field test" version of the PDDBI covering its development and psychometric characteristics was published in 2003 (Cohen, et al., 2003a, 2003b) with the updated professional version published in 2005 (Cohen and Sudhalter, 2005). **Tables covering the adolescent years from 12 years, 6 months to age 18 years, 5 months were published in 2017 (Cohen, 2017). These age ranges will be discussed below regarding the most recent 2020 Annual Report to Congress (# VI).**

As shown in Table 1, there are seven domains and two overall composite scores within the Approach-Withdrawal Dimension, three domains and two overall composite scores within the Receptive/Expressive Social Communication Abilities dimension, as well as an Autism Composite score (comprising 6 of the 10 domains) measuring the overall severity of the core signs of ASD (socialization and communication deficits and repetitive/ritualistic behaviors) in a statistical sense (see below). These domains were included to enable therapists to gauge the efficacy of their interventions by comparing the magnitude of changes in those T-scores **that are relevant to the specific behaviors targeted for change.** For example, if the focus of intervention were receptive language skills, then the therapist would examine changes in the Learning, Memory and Receptive Language Domain. If the focus were on decreasing aggressiveness toward self or others, then the therapist would examine changes in the Aggression Domain. Neither of these domains is part of the overall Autism Composite score. Unfortunately, the DoD reports described below exclusively examine only the Autism Composite on the assumption that all interventions target decreases in core signs of ASD (i.e., socialization/communication problems and repetitive/ritualistic behaviors; more on this below).

The PDDBI was standardized, separately, on data from parents and teachers/other professionals about children with documented autism resulting in two different forms (PDDBI-P/X for parents and PDDBI-T/X for teachers) as noted above. Agreement between these two types of informants varies with the domain. In general, agreement is much better for the Receptive/Expressive Social Communication Abilities dimension than the Approach-Withdrawal Dimension or Autism Composite. **This is**

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

not unusual and is characteristic of other similar rating scales (Lane et al., 2013). Parents and teachers see the children in different situations with different expectations and rules for managing problem behaviors and often make judgments relative to their experience with other children. The PDDBI manual provides tables for assessing degree of discrepancies between parent and teacher forms for each domain and this information can often be useful to the clinician or therapist (e.g., in diagnosing separation problems or in assessing degree of generalization of skills acquired in different settings).

Why is the normal distribution of the PDDBI T-scores important?

As noted, the scores on the PDDBI are normally distributed meaning that the distribution of scores across all cases with ASD follows a "bell-shaped" curve with most scoring in the "average" range with extreme scores relatively rare. In the general population, IQ scores follow such a pattern and are standardized such that most people will have an IQ of 100 +/- 15 points (where 15 is the "standard deviation" or SD – a measure of the variability in the scores). Very few will have IQs above 115 or below 85. The PDDBI follows a similar pattern except that the average "T-score" for children with ASD is set at 50 (with an SD of 10 points). All the domains and composite T-scores of the PDDBI are normally distributed (i.e., the overall distribution of T-scores has a "bell-shaped" curve). This is important for several reasons.

First, the normal distribution has certain mathematical properties that are useful in describing the ASD cases. As I noted above, the average PDDBI T-score is 50 points with an SD of 10 points. Given what we know about normal distributions, this means that 68% of all children with ASD will have scores that range between 40 and 60 and 95% will score between 30 and 70. Those cases with ASD that fall outside these ranges are, by definition, atypical for ASD. Many of the cases with scores less than 30 or 40 either have very mild features of ASD or do not actually have ASD (see Autism Diagnostic Observation Schedule-Generic (ADOS-G; Lord et al., 2000) results in the PDDBI Manual, Cohen and Sudhalter, 2005, pp. 74-77) while those with scores greater than 60 or 70 are typically very severely affected and are more likely to suffer from intellectual disability and/or co-occurring psychological issues (see "Albert" case in the PDDBI manual, Cohen and Sudhalter, 2005, pp. 30-32 and Cohen et al., 2010). This issue will come up later in our discussion of the outcome data presented by the DoD in their reports.

Second, the normality of the T-scores enables sophisticated statistical analyses to be carried out on these scores since almost all such models assume that the data to be analyzed are normally distributed.

Lastly, the normality of the distributions makes certain computations more sensitive to interpretation of group differences. One such measure is the "effect size", a measure of the importance of the results (more on this below).

Critiques

I. Report to Congress: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress Fourth Quarter, Fiscal Year 2018

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

In this report, the DoD states that it is tasked with reporting "the health-related outcomes for beneficiaries under the program." (p. 2). On p. 9, they note that, to this end, three outcome measures are required, one of which is the PDDBI and is the only one of the three for which reports are generated every 6 months. I would note that in this report and in all subsequent reports which this analysis reviews, the name of the PDDBI is consistently misstated. It is called the PDD Behavior Inventory, not the Pervasive Developmental **Disabilities** Behavior Inventory. I bring this up only because it suggests that the DoD officials who wrote the reports have not actually seen the PDDBI or the PDDBI Manual or are unfamiliar with Autism diagnostic nomenclature where PDD means Pervasive Developmental **Disorder**.

There are several problems with how the PDDBI is described and analyzed in these reports which make the Autism Composite Score data suspect.

Scoring Problems

On p 9, it is noted that this is the first quarter in which two sets of PDDBI scores were available for comparison. The DoD state that of 16,277 beneficiaries enrolled, about 14,000 had at least one outcome measure and, of these, 1,577 had "usable" scores. They then say:

> "Many beneficiary scores noted "0", indicating an incomplete or an unable to answer section of the PDDBI based on a variety of factors (i.e., direction to not complete a section of the child is non-verbal. Additionally, this number ... represents beneficiaries from only the west region, as the current contract requirements did not specify reporting scores by outcome measure form type (parents vs, teacher form)."

Comments:

1. **The reason for the elimination of most of the data, i.e., scores of "0" is factually incorrect as noted above.** *A score of "0" means that the behavior being rated is not present. It does not mean that the item is missing*. I can only conclude that the persons responsible for scoring the PDDBI had not bothered to read the manual. If we take this description to its logical conclusion it would imply that forms for which the child was reported to not show a behavior were not included in the analysis thereby invalidating the interpretation of the results **since it is possible that cases who had improved (e.g., no longer showing repetitive language or hand flapping) would not be included in the follow-up results, likely biasing the data toward cases that did not change because of the significant number of discarded forms. Until these "missing" data are rescored and included, all of the analyses are suspect.**

2. **There is no indication that, of the forms deemed "usable", the person completing the form at baseline was the same person completing the form on follow-up, which is essential.** Indeed, **there was no information about the informants including whether or not English was their primary language**. Usually, the parent completing the form is the mother, but this is not always the case and could be even more of an issue for active duty family members. For such cases, it would be beneficial to have both parents independently complete forms at each assessment. Failure to ensure identical informants means that any differences due to time were confounded with the perceptions of the different raters. **This therefore introduces additional "noise" into the**

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

data and would likely result in much greater variability in measures of change over time. In turn, this would diminish the sensitivity of statistical tests to actual change in T-scores over time.

3. **There is no indication as to who these children are in terms of their demographics.** There is no mention of their ages, sex, IQ, family or health issues (e.g., having another affected child in the family, marital discord, child health). It is known that younger children often respond better to intervention for a variety of reasons. Interventions in older children with intellectual disability often tend to focus primarily on self-care skills and behavior management issues, not so much on core signs of ASD.

4. **All of the above imply that this is a "garbage in, garbage out" (GIGO) situation to use a computer science acronym. Faulty input equals faulty output, or, in this case, faulty data equals faulty and unsustainable conclusions.**

On page 10, the DoD present Autism Composite score differences for the group as a whole and show the distribution of these differences divided into three outcomes defined by a 10 point difference in the T-scores in Figure 1. They state:

> "... approximately 87 percent (1,365) of beneficiaries made little to no change in their symptom presentation after six months of ABA services (647 or 41 percent of the population had no change in PDDBI – parent score; 718 or 45 percent had less than one Standard Deviation (SD) change in PDDBI – parent score). Of significance, six percent of the population had a decline of one SD or more indicating worsening symptom presentation after six months of ABA services. Only seven percent of the sample had improvement (1 SD or better) in symptom presentation after six months of ABA services."

Comments:

1. These statements by the DoD are unsupportable because the data are largely uninterpretable for the GIGO reasons specified above.

2. The data that are examined are analyzed only in a descriptive manner and make faulty assumptions regarding the magnitude of expected change over time, requiring an unusually high effect size (cf. point 5 below). I did not see any inferential statistics applied to these data (e.g., a paired-sample t-test or repeated measures ANOVA) which would have been a more appropriate form of analysis, also noted in point 5.

3. Only the Autism Composite T-score is analyzed on the assumption that interventions focus exclusively on decreasing core symptoms of ASD (e.g., social communication deficits and repetitive/ritualistic behaviors). While this is likely to be the case in younger children based on reports in the literature, it is less likely the case in older children where development of life skills (e.g., self-care, use of mass transit, etc.), use of augmented communication devices in non-verbal cases, and behavior management issues are often target goals.

4. Instead of examining only the Autism Composite T-score, there should have been a breakdown by age as well as by the various targets of intervention pertinent to a given child with

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

those targets matched to relevant PDDBI domains (e.g., reductions in rituals, decreases in repetitive language, decreases in aggression, improvements in social interaction, etc.).

5. The methods of statistical analysis of the data appear to be purely descriptive and not consistent with analyses of outcomes reported in the literature. For a simple before and after measure of change, a more appropriate starting analysis is a paired sample T-test where the DoD would then present the results based on the means (averages) and standard deviations (SDs) at times one and two for all domains and composite scores of the PDDBI, along with effect size measures for each of the domains and composite scores, taking the correlation between the two measures into account. Measures that are highly correlated over time (such as the PDDBI domains and composite scores) are more sensitive to change and require much smaller differences to be statistically and clinically significant. Effect size measures the actual magnitude of change which is, in turn, related to clinical relevance, and this is discussed in more detail in my comments on the most recent 2020 Annual Report to Congress. The assumption that scores must change by 10 points (one SD) is a very stringent requirement and much greater than needed to demonstrate an effect. If the DoD still wishes to analyze only change scores, then a covariance analysis taking baseline levels into account would be needed since there is more "room" for improvement in those who are more severely affected. Such an analysis would take baseline severity into account. **It should be noted that in the PDDBI Manual, the 90 percent confidence interval for the Autism Composite T-score is not 10 points but, instead, ranges from 4 to 6 points, depending on the age of the child. This means that 90 percent of the time, repeat assessment for a given child's score should vary by no more than 4 to 6 points, about one-half of the DoD requirement.**

6. **With respect to point 5, the DoD does not present the actual means and SDs of the Autism Composite T-scores at the two different time points, along with their correlation, making it impossible to gauge the actual magnitude of the effect with this problematic data.**

7. More appropriate statistical analyses that would take into account the various issues discussed above include repeated measures analyses of variance or covariance and repeated measures multivariate analyses of variance and covariance wherein relevant factors such as age, sex, goal plans, degree of impairment, program stability, health status (which could account for some cases of worsening of behaviors), etc. could be analyzed for their effects on ABA program relevant PDDBI domain and composite scores. This would assume that the above-mentioned data scoring issues have been solved.

On page 11, the DoD notes:

"Also reviewed was the concordance/discordance between parent and teacher or Board Certified Behavior Analyst (BCBA) completed forms….Approximately 60 percent of the completed parent and teacher forms were within 10 points or one SD….Of the remaining 40 percent where there was greater difference in scores (1 SD or greater), 70 percent of the parents scored worse symptom presentation than the teacher or BCBA. According to the research regarding the PDDBI, there is a high degree of interrater reliability between the parent and teacher forms. This discrepancy in TRICARE beneficiaries requires further exploration."

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Comments:

1. The same GIGO issues regarding the quality of the data obtained pertain here as well: **A) The reason for the elimination of most of the data, i.e., scores of "0" is factually incorrect. A score of "0" means that the behavior being rated is not present. It does not mean that the item is missing. B) There is no indication that, of the forms deemed "usable", the person completing the form at baseline was the same person completing the form on follow-up, which is essential and C) there is no indication as to who these cases are.**

2. The more obvious issue, however, is the fact that this quote clearly indicates that the DoD authors did not bother to read the PDDBI Manual. **It is _not_ the case that the inter-rater reliability between parents and teachers is high for the Autism Composite Score.** In the PDDBI Manual, Table 5.9, page 57, it is indicated that the correlation between parents and teachers for the Autism Composite Score is 0.32, **which is not "high"**. Correlations vary between -1.0 (perfect negative relationship) to +1.0 (perfect positive relationship).

3. The reasons for this are discussed above and, as noted, the PDDBI manual provides tables for the clinician regarding the degree to which lack of agreement between parents and teachers is significant and needs to be examined for its clinical relevance.

II. Report to the Committees on Armed Services: The Department of Defense Comprehensive Autism Care Demonstration Annual Report June 2019

This report is directed to the Committees on Armed Services and is based on 2018 data and, with respect to the PDDBI, covers the same data discussed in detail above. This report summarizes the conclusions drawn in the previous report for the 1,577 beneficiaries in the West region.

On page 17, the DoD state:

> "The first analysis of the PDDBI found the majority of TRICARE beneficiaries (86 percent) had little to no change in symptom presentation over the course of 6 months of ABA services. It may be that more time is required to see change; however, input on treatment progress should be collected in short intervals so that time does not pass with ineffective treatment. Additionally, the 40 percent discrepancy in responses between parents and teacher/BCBA is also of note suggesting the DHA should explore the possible reasons for the wide range in perceptions of symptom presentation, to include evaluating the utility of the parent form and of this measure generally. Further analysis is required to observe trends and utility. While it is concerning that 86 percent of the population saw little to no change, the MCSCs (with government oversight) will work with the providers to ensure effective treatment is being delivered."

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Comments:

1. With respect to the "little to no change" over 6 months, my comment simply is that the data, as presented, are uninterpretable for the GIGO reasons specified above: **A) The reason for the elimination of most of the data, i.e., scores of "0" is factually incorrect. A score of "0" means that the behavior being rated is not present. It does not mean that the item is missing. B) There is no indication that, of the forms deemed "usable", the person completing the form at baseline was the same person completing the form on follow-up, which is essential and C**) **there is no indication as to who these cases are.** Further, there is no basis for these statements as the DoD has not defined up-front, what they consider to be clinically relevant improvement and, instead, falsely assume that T-score changes have to be 10 points or greater for an improvement to be noted. I will discuss this further in my comments on the most recent report to Congress (numbered VI).

2. The agreement between parent and teacher issue has been discussed above but, again, repeats a false narrative regarding the purported high level of concordance between these types of informants on the Autism Composite T-score. As noted, the PDDBI manual clearly indicates that the agreement is not high for the Autism Composite T-score but is high for domains and composite scores that measure social and language competence. **It is clear that the Department of Defense authors have not bothered to read the manual.**

III. Report to Congress: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress Second Quarter, Fiscal Year 2019

This quarterly report to congress provides additional 12-month follow-up data on the PDDBI on the same 1,577 cases. The errors noted above still pertain for this quarterly report. They state (p. 9):

> "The same 1,577 beneficiaries submitted in the July to September 2018 quarterly report were reviewed and of the 1,577 beneficiaries, 709 beneficiaries had usable scores for comparison of the PDDBI. Many beneficiary scores noted "0," indicating an incomplete or an unable to answer sections of the PDDBI based on a variety of factors (i.e., direction to not complete a section if the child is non-verbal)."

Comments:

As noted, this is a factually incorrect statement which, in this report, indicates that only 45% of the 1,577 cases were analyzed here because of this wrong assumption. This is a change from the previous two reports which examined 1,577 cases suggesting inconsistencies that make it difficult to compare outcomes across reports.

> "For the reporting periods of July to September of 2018 and January to March of 2019, based on the Autism Composite Score on the parent form of the PDDBI (which is a measure of lack of appropriate social communication skills along with repetitive/ritualistic behaviors),

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

approximately 81 percent (572 total comparable Parent Forms) of beneficiaries made little to no change in their symptom presentation after six months of ABA services). Of significance, seven percent of the population had a decline of one standard deviation (SD) or more, indicating worsening symptom presentation after six months of ABA services. Only 12 percent of the sample had improvements (one SD or better) in symptom presentation after 6 months of ABA services. See Figure 1 for the distribution of change scores for the parent score for the 6-month comparison period." (p. 9)

Comments:

As I noted above, these faulty data are examined using a difference score measure for the Autism Composite T-score which assumes that core signs of ASD are the targets of intervention for all cases irrespective of age or functioning level and assumes that meaningful change means a difference in T-scores of 10 points or more. As well, the actual means and SDs at the different time points are again unspecified and the actual methods of analysis of the data are also unspecified.

"For the reporting periods of January to March of 2018 and January to March of 2019, based on the Autism Composite Score on the parent form of the PDDBI, approximately 76 percent of beneficiaries made little to no change in their symptom presentation after 1 year of ABA services. In this comparison, nine percent of the population had a decline of one SD or more, indicating worsening symptom presentation after one year of ABA services. Only 16 percent of the sample had improvements (one SD or better) in symptom presentation after one year of ABA services. Of particular interest, Figure 2 indicates a 4 percent increase in improved ASD symptoms scores and a 2 percent increase in worsening ASD symptoms scores after 12 months of ABA services. Of concern is that 76 percent of beneficiaries continue to not report symptom improvement after 12 months of ABA services." (p. 10).

Comments:

My comments here are the same. **These data are not only scored incorrectly but they are analyzed only in a descriptive manner rendering interpretation meaningless, as noted in my comments above**. Given that there are now three time points, a more appropriate analysis would be a repeated measures analysis of variance on the Autism Composite T-score, **assuming that the data were scored correctly**, with tables and figures showing the actual means and SDs (or, even better, the confidence intervals around the means since the analysis would be concerned with average scores) at each time point. **However, the same faulty assumptions plague these data resulting in additional and significant loss of data. Of concern is that the elimination of cases who scored "0" on relevant items could indicate inadvertent selective elimination of children who improved over time again, a GIGO problem.**

"Also reviewed was the concordance/discordance between parent and teacher (or Board Certified Behavior Analysts (BCBA) completed forms of this quarter's score submission. Of the 709 beneficiaries pulled for this analysis, 647 beneficiaries had both parent and teacher forms submitted for this reporting quarter. Approximately 58 percent of the completed parent and teacher forms were within 10 points or one SD of one another suggesting that there was agreement in slightly more than half of the T-scores for the Autism Composite Score regarding the perception of symptom presentation. According to the research regarding the PDDBI, there is a high degree of interrater reliability between parent and teacher forms. This

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

discrepancy in TRICARE beneficiaries continues to require further exploration. See Figure 3 for the distribution of scores for the parent form and Figure 4 for the teacher form." (p. 11)

Comments:

This is addressed above. I simply repeat the statement that Autism Composite T-scores for parents and teachers do not have a high level of agreement indicating that the DoD are repeating an error noted above and indicating their failure to read the manual. The fact that, for these data, the distribution of teacher Autism Composite T-scores is shifted toward the less severe range and, overall, lower than parent scores is not surprising given that many children with ASD tend to respond well to highly structured and predictable environments. However, this effect is often dependent on age, functioning level, the home environment, and presence or absence of co-occurring psychological disorders such as attention deficit disorder and/or various anxiety disorders, both of which are common in this population and not discussed or included in the DoD analysis.

"Determining health-related outcomes is an important requirement added to the ACD. A contract modification, effective January 1, 2017, provided direction for MCSCs to begin collecting the outcome measures data for all ACD participants. The MCSCs use these scores, as well as other scores and data, to guide and engage ABA providers in identifying treatment plan development and adjustments that may be required to see improvements. The DHA remains committed to ensuring all TRICARE-eligible beneficiaries diagnosed with ASD reach their maximum potential, and that all treatment and services provided support this goal. Based on this reporting quarter outcome measures data, the majority of TRICARE beneficiaries (76 percent – parent form) had little to no change in symptom presentation over the course of 12 months of ABA services, with an additional nine percent demonstrating worsening symptoms. Additionally, the 42 percent discrepancy in responses between parents and teacher/BCBA (Figure 3 and 4) is also of note, suggesting DHA should continue to explore the possible reasons for the wide range in perceptions of symptom presentation. Further analysis is required to observe trends and utility. While it is concerning that 76 percent of the population saw little to no change, the DHA via the MCSCs will work with the providers to ensure effective treatment is being delivered." (p. 12).

Comments:

While it is good that they are using data to inform ABA providers in identifying and improving treatment plans, the analyses presented here will not achieve these goals.

1. **As noted, it is incumbent upon persons administering the PDDBI to score it properly so that scores are interpretable which, in turn, requires these people and those analyzing the data to read and understand the PDDBI Manual (and this is explicitly stated on page 5 of the PDDBI Manual).** GIGO issues persist in this Report to Congress, thus the conclusions drawn in this statement are not justified.

2. **Persons completing the forms at baseline and follow-up points should be the same** so as not to confound change over time with different rater biases and perceptions. It is unclear if this has occurred.

3. **Exclusive focus on the Autism Composite T-score as an outcome measure assumes that all**

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

cases with ASD, irrespective of age and functioning level, have, as ABA goals, reduction in core signs of ASD. This is <u>not</u> necessarily the case, <u>especially for older children and those with severe intellectual disability</u>. Instead, **there should be a matching between the individual treatment plans for a given child, and changes expected to be seen on relevant PDDBI domains and composite scores.** For example, interventions targeted toward decreases in anxiety and irritability would be measured by changes in the Fears and Aggression domain T-scores while interventions focusing on increasing social skills would be measured by changes in the Social Approach domain T-score. Agreement between parent and teacher informants here would serve as a measure of generalization across persons and environments. Any significant improvements in other domains or composite T-scores could indicate unintended generalization across behaviors, a good sign. Conversely, increases in domains assessing challenging behaviors would be an indication of a need to either modify the treatment plan and/or have the child examined for health problems that are commonly seen in ASD and that could increase such behaviors (cf. Restrepo, et al., 2020).

IV. Report to Congress: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to the Armed Services Committees, Third Quarter, Fiscal Year 2019

This third quarter report is to the Armed Services Committee. DoD presents PDDBI data on the same three time periods noted for the **Quarterly Report to Congress Second Quarter** (roman numeral III above) and they repeat the same faulty assumptions and unsupportable conclusions in that report. The comments in this report on the PDDBI results are as follows:

> "For the reporting periods of April to June of 2018 and April to June of 2019, based on the Autism Composite Score on the Parent Form of the PDDBI, approximately 70 percent of beneficiaries made little to no change in their symptom presentation after 1 year of ABA services. In this comparison, 10 percent of the population had a decline of one SD or more indicating worsening symptom presentation after 1 year of ABA services. Only 20 percent of the sample had improvements (1 SD or better) in symptom presentation after 1 year of ABA services. *While the improved scores have increased as compared to the 6 month comparison, it is important to note that the sample size for matching scores for this comparison was only 240, and, therefore it would be difficult to make any conclusions regarding overall change.* (italics added by Cohen) The DHA is further analyzing why the reduction in matching sample scores occurred over the year and plans to implement process improvements for collecting outcomes measures in the next manual revision." (p. 10).

Comments:

Here, the DoD repeats the same problematic assertions noted above: A) the reason for the elimination of most of the data, i.e., scores of "0" is factually incorrect. A score of "0" means that the behavior being rated is not present and does not mean that the item is missing. B) There is no indication that, of the forms deemed "usable", the person completing the form at baseline was the same person completing

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

the form on follow-up, which is essential, and C) there is no indication as to who these cases are, demographically. They now add a further caveat to the third time-period. For some reason, the number of cases with PDDBI data is no longer 572 but is now 240 and the DoD indicate that they do not know why. They also assert that this reduction in sample size makes it difficult to interpret change. I would simply note here that this further reduction in the sample by over 50% is problematic and inconsistent with previous reports. Further, I would note that, using their own criterion of a one standard deviation improvement over time (10 points – a very large effect size – see below for my comments on the most recent report to congress), twenty percent of their reduced sample achieved this huge effect but they minimize this effect by punting to problems with the sample size (which most researchers would consider quite large). As before, their data analysis method is not described and basically consists of scoring distributions of difference scores without specifying what these are based on and with no information on the cases, their families, ages, etc. More appropriate data analyses are described above.

V. Report to Armed Services Committees: The Department of Defense Comprehensive Autism Care Demonstration Quarterly Report to Congress, Fiscal Year 2019

This quarterly report to the Armed Services Committees is brief and summarizes previously reported analyses of PDDBI data. On page 2, they note the following:

"…while outcome measures are required for this report, the Department believes that a more in-depth review of records is necessary at this time as the lack of significant progress over the last three reporting periods has led to questions and concerns about the appropriateness and/or effectiveness of services under the ACD that require further analysis. A more detailed analysis will be available in the next annual report."

Comments:

The previous reports summarizing outcome with the PDDBI were faulty for all of the reasons cited above leading to the incorrect conclusion of a "lack of significant progress." No such conclusions can be drawn from these previous analyses. One cannot conclude that there was no significant progress when:

1. The data were scored incorrectly leading to initially deleting 45% of cases that may have improved and then deleting an additional 50% of the remaining cases which throws into question the relevance of any DoD analysis and conclusions in this report.

2. There was no assurance that all measures of outcome were completed by the same person.

3. There was no demographic and health-related information on the cases themselves.

4. The data analyses that were performed focused only on the Autism Composite T-score on the assumption that <u>all</u> ABA interventions were trying to decrease core symptoms of ASD (i.e., social and language deficits and repetitive or ritual-

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

istic behaviors) irrespective of age, level of cognitive or behavioral impairment, etc.

5. The data analyses were not specified and were descriptive in nature but did not provide basic data on actual means and SDs at each time point. Instead, they assumed an extremely large effect size for this single measure of autism severity as a measure of improvement without considering other important factors such as age, intellectual disability level, the goals of intervention and initial severity of ASD.

6. The points raised above are also recognized on page 10, last paragraph, of this report, as a result of concerns raised by advocacy and provider groups and promises a re-evaluation in the July 1, 2020 report.

VI. Report to the Committees on Armed Services of the Senate and House of Representatives: The Department of Defense Comprehensive Autism Care Demonstration Annual Report 2020

In this most recent report to the United States Congress Committees on Armed Services, additional analyses were performed on the PDDBI data covering cases receiving as much as 18 months of ABA services because of "letters of concern" regarding the previous findings (p. 17). Here, data are presented from both East and West regions for cases having at least 18 months of services since January 1, 2018. Cases not having baseline, 12- and 18-months of data for the PDDBI Parent form were not included. The total numbers of cases have now increased from previous reports and total 3,794.

Comments:

1. It is not clear why the number of cases increased from 240 to 3,794. This is not explained.

2. It is also unclear if the misinterpretations reflected in the previous 5 reports to Congress of the scoring of the PDDBI were corrected.

3. It is unclear if there was an attempt to ensure that the informants at the three time points were the same.

On Page 17, the following statement is made:

"Both regions demonstrate similar results across all figures thus increasing the confidence of the data. Please note that many of these beneficiaries received more than just ABA services, therefore, it is impossible to know for certain whether the changes reported here are due to ABA services, other services, maturation of the individual, or a combination of factors."

Comments:

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

1. As I understand the directive, the reason for gathering PDDBI data and comparing results over time is to document whether improvement is or is not seen, not what the reason is for any changes. Indeed, as noted, the agency is not carrying out a double-blind, randomized control trial so causation at the level of the group as a whole cannot be inferred.

2. With respect to "maturation", which I will also address later, this cannot be an explanation due to the design of the PDDBI. The evidence for spontaneous improvement in most cases with ASD is meager. Having said that, the PDDBI manual shows that "raw scores" for social and language skills are better in older children relative to younger children. **However, the T-scores are computed based on the age of the child. If the children were to improve in these areas simply due to maturation, the T-scores would stay the same.**

Starting on page 18, the DoD provides more analyses of their follow-up data using the same methods described above which I have noted to be inappropriate for the nature of the data. Given this caveat, I will now address various concerns I have with the results as presented. These concerns are in addition to the data quality issues described above.

P. 18. It is stated that:

"The histogram in Figure 4 represents a comparison for both West and East regions at 12 and 18 months for the beneficiary actual point score change (point score change is defined as the difference between baseline score and the 12 or 18 month score). Of note, a decrease in score on the PDDBI denotes improvement in symptom presentation. After 12 months of ABA services, West region beneficiaries had an average change score of -1.22, and East region beneficiaries had an average change score of -1.93 with a 95 percent confidence interval (CI) that accounts for random error, meaning that the true score falls within the range of -1.76, -0.68 and -2.38, -1.48, (West, East respectively). After 18 months of ABA services, West region beneficiaries had an average change score of -2.50, and East region beneficiaries had an average change score of -2.40, with a 95 percent CI that accounts for random error, meaning that the true score falls within the range of -3.03, -1.98 and -2.89, -1.91, (West, East respectively). For both 12 and 18 months, the p-value, which is an indicator if there is a difference from zero, is less than 0.05 which means that this data is statistically significant. Although there was a statistical difference in scores at 12 and 18 month, the presented changes do not necessarily indicate clinical improvement especially since the gains are extremely small. To our knowledge, there is no available literature defining how much change would be considered clinically significant…"

Comments:

1. As I have noted, these analyses are problematic. There is no mention as to the statistical model used to analyze the data and, again, there is no mention of the actual means and standard deviations for baseline and follow-up reports and no description of the sample on which the data are based. Only the difference scores are presented, as before, and we are told that the actual differences are quite small, changing about 2 points at 12 months and about 2 to 3 points at 18 months. The p value (an indication of the likelihood that the change score is due to chance)

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

is cited as less than 0.05 (i.e., less than 5%). Yet, the p values shown in Figure 4 are orders of magnitude less than this, e.g., the East region p value is less than 0.00000000000000010! These effects, even given the problems with the data cited above, appear to be highly statistically significant indicating a rather strong "signal" of an effect with this problematic data and are greater over time. Yet the significance of this effect is minimized by the DoD. Are these differences "extremely small"? This depends on the measure and how much of a change is considered clinically significant. The DoD should have computed what is called the "effect size", a measure of the meaningfulness of the change. This computation could be easily computed if the DoD had analyzed and presented the data properly, e.g., at the most elementary level, using a paired-sample t-test comparing two different time points (cf. https://www.real-statistics.com/students-t-distribution/paired-sample-t-test/cohens-d-paired-samples/).

2. **In terms of the DoD's lack of awareness of the clinical meaningfulness of the size of the Autism Composite T-score, this is addressed directly in the validity section of the PDDBI manual.** For example, Figures 5.3 and 5.4 from the manual show the different profiles for children grouped according to their Vineland Adaptive Behavior Scale (VABS, a measure of social, language, daily living, and motor skills) score ranges (adequate to moderately low, mild to moderate, and severe to profound, where all ranges are relative to typically developing persons) and their Autism Diagnostic Observation Scale-Generic (ADOS-G) ranges (autism, spectrum, not spectrum). The Autism Composite is the right-most label along the X-axis. As shown, the average difference between a child classified as having Autism vs. the milder Spectrum classification is about 5 points on the PDDBI Autism Composite T-score. Similarly, the average difference between children in the severe to profound range and the moderate to mild range on the VABS is about 4 points for the Autism Composite. Given that the average change reported for these data is about 3 points at 18 months it can be inferred that, as a group,

Figure 5.4

Multivariate F (28,400) = 5.75 p < 0,0001

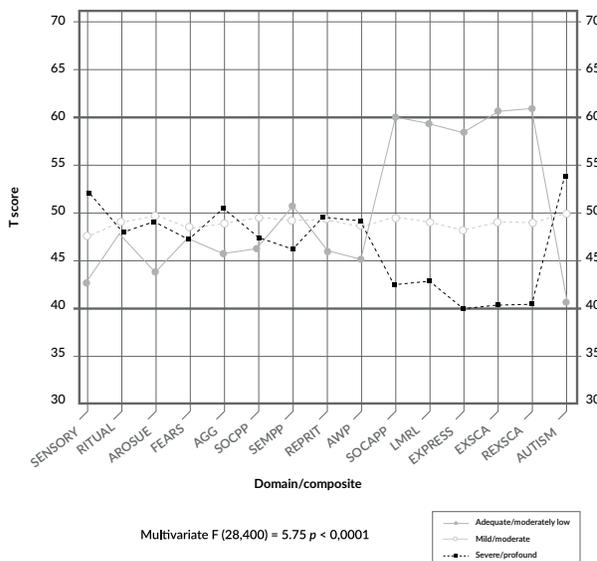Legend: Adequate/moderately low; Mild/moderate; Severe/profound

Figure 5.4 Profile of the mean PDDBI-PX/PDDBI-P domain and composite scores for each of the combined Vineland groups, with significance assessed with a multivariate, repeated measures analysis

Figure 5.3

Multivariate F (28,400) = 5.75 p < 0,0001

Legend: Adequate/moderately low; Mild/moderate; Severe/profound
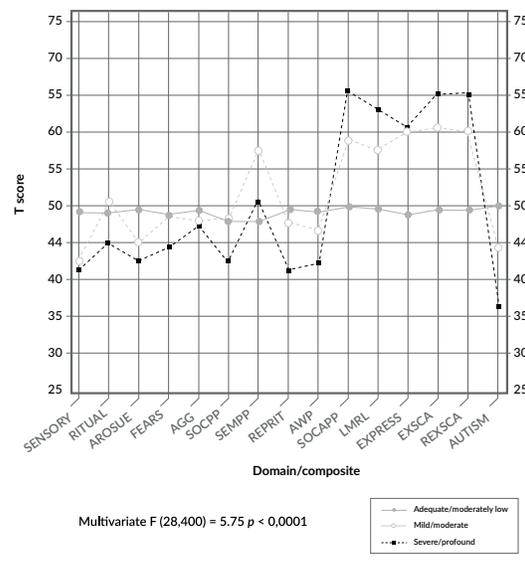
Figure 5.3 Profile of the mean PDDBI-PX/PDDBI-P domain and composite scores for each of the ADOS-G groups, with significance assessed with a multivariate, repeated measures analysis.

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

the children were actually moving toward a less severe form of autism with likely improvement in their adaptive behavior skills over time. **T**hese figures are reproduced here:

On p. 19, the DoD states:

"The remaining outcome measures data reported in this report are analyses that focus on beneficiaries' outcomes after 18 months of ABA services. Figure 5 depicts the average percent of change scores for beneficiaries by their baseline score. Figure 5 is different from Figure 4 in that Figure 5 is a percent change not a point change. For both West and East regions, all groups with scores 40 and above demonstrated statistically significant change on the PDDBI. However, those beneficiaries with the most severe baseline scores (80-100) demonstrated the greatest change in PDDBI score after 18 months. Additionally, scores 39 and below demonstrated no change from baseline score. Again, it is unclear if ABA services were the change agent impacting these percent score changes or if another variable, or combination of variables, created the change. Additionally it is unclear if any of the change is of clinical significance."

J Autism Dev Disord

Table 4 ROC results for best composite measures and discrepancy scores

	Sensitivity	Specificity	Cut-off	AUC	AUC 95% CI	PPV	NPV
Austism versus NS							
EXSCA/C	97	90	66	96	91-100	92	96
REXSC/C	97	90	66	96	92-100	92	96
AUTISM/C	91	90	42	94	87-100	91	90
SOCDSC	100	90	24	96	91-100	92	100
SEMDSC	100	73	13	87	76-97	81	100
PDD-NOS versus NS							
EXSCA/C	94	86	69	92	83-100	80	96
REXSC/C	100	79	73	92	84-100	74	100
AUTISM/C	100	79	32	90	80-99	74	100
SOCDSC	100	89	26	94	86-100	84	100
SEM DSC	97	78	13	86	74-97	72	98
Autism/PDD-NOS versus NS (entire sample/PQ matched pairs - see text)							
EXSCA/C	88/95	93/85	67/69	95/94	89-100/86/100	96/86	82/94
REXSC/C	96/90	85/85	69/69	95/94	90-100/86/100	92/86	93/89
AUTISM/C	100/90	79/95	32/42	94/97	87-100/93/100	85/95	100/90
SOCDSC	98/100	93/95	25/25	96/98	90-100/94/100	96/95	96/100
SEMDSC	98/95	79/80	13/15	88/88	79-98/76-100	88/83	81/100

All scores (except for cut-off scores) are in percent; AUC = area under the ROC curve; PPV and NVP = positive and negative predictive validity

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Comments:

1. Here, the DoD examines Autism Composite T-scores broken down by initial severity level using percentage change from baseline as the measure instead of the absolute difference score. They note that the greater the initial severity, the greater the percentage change ranging from little to none for the cases scoring less than 40 but increasingly larger differences for the more severely affected cases. These effects are interesting and not surprising since, as I noted above in the initial description of the normal distribution of PDDBI T-scores, very few cases with autism score less than 40 and, in published data on cut-off scores for the PDDBI, for children who are matched on IQ, a score of 42 on the Autism Composite T-score best separates those with ASD from those without ASD (see Table 4, above, for the Autism/C - "cut-off" score column, third line from the bottom, from Cohen et al., 2010). Therefore, in contradiction to the DoD conclusions, the data presented in this last report suggest that cases more likely to have ASD are improving in severity by 18 months and the greater the impairment, the better they do, for all cases irrespective of age, quality of the data, health status, etc., yet the DoD completely mischaracterizes these results.

2. The clinical significance of these results can easily be found by consulting the PDDBI manual, as cited above.

3. The statistical analyses used for these data are still not described and appear purely descriptive in nature.

On page 20, the DoD state:

"Figure 6 analyzes the data by age and gender. As seen in both data sets from the West and East regions, all ages groups except the East region ages 15-20 years demonstrated a statistically significant percent change from baseline scores. Additionally, beneficiaries who are younger, ages 0-5 years, had a percent change that was statistically significantly greater than the other age ranges, meaning the younger beneficiaries had a greater level of percent change from their baseline scores. All other age ranges had a CI that overlapped, indicating there was no statistically significant difference between age groups above 5 years."

Comments:

1. The figure they refer to, Figure 6, examines the average change from baseline for four different age groups: 0-5, 6-10, 11-15, and 15-20 years. The PDDBI likely used for these analyses is the original one which covers ages from 18 months to 12 years, 5 months. How were scores computed for the 1-year old infants and those children over 12 years, 5 months? The adolescent tables for the PDDBI were published in 2017 but do not cover cases over 18 years, 5 months. Were the latter cases scored using these new tables? This discrepancy needs to be clarified.

2. The percentage change in the severity of core signs of ASD was greatest in the youngest children. This is not surprising given the fact that, at this age, the goals of intervention are usually targeted toward ameliorating the severity of these signs by teaching

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

attending, language, play, and social interactive skills and given results from the early intervention literature. As I noted above, goals for older children and those with greater intellectual impairment may or may not target these core signs. Further, changing behaviors in older individuals is often more difficult because they have had more time to experience failure and to have learned, over many years, successful but problematic ways of dealing with others and their environment (e.g., through aggression or self-injury).

3. The exclusive focus on the Autism Composite T-score as a measure of response to intervention may partially explain these results. There are 10 different domain scores and 4 other composite scores that should be examined here. These, in turn, should be examined with respect to the goals for a given child, e.g., decreasing challenging behaviors, decreasing rituals, improving language skills, etc. and monitored for change, using appropriate statistical models.

On page 21, the DoD state:

"Figure 7 depicts the percent change in baseline PDDBI PACS (*note: the DoD uses this to refer to the Parent Autism Composite Score – Cohen*) compared to the total number of hours of rendered ABA services over 18 Months. This number is the total number of direct one-to-one hours of paid claims for Category III Current Procedural Terminology (CPT) code 0364T/0365T and Category I CPT code 97153 (Adaptive Behavior Treatment by Protocol). For the East region, the trend line indicates that beneficiary scores worsened with more hours of ABA services. In the West region, there is no statistically significant correlation between the total number of direct hours rendered and outcome measure scores. The West region trend line demonstrated a flat trend line noting no correlation with rendered hours of ABA services. There does not appear to be a correlation between outcome measures and the number of hours rendered. In other words, the number of hours rendered does not appear to impact outcomes. If the amount of direct ABA services was correlated with improvement, the trend line would demonstrate a statistically significant negative slope. Therefore, any percent change in PAC scores over time (Figure 4) cannot be directly attributed to hours of ABA services provided under the ACD, and could be due to other factors such as developmental growth/maturation and/or other concurrent treatment."

Comments:

1. It is impossible to draw conclusions from the lack of correlation between the Autism Composite T-score and number of hours of intervention without knowing more details such as when intervention started, family circumstances, goals of intervention and so forth and without knowing whether the PDDBI forms were scored correctly and completed by the same person at the different time points.

2. The supposition that any change observed can be attributed to another intervention is pure speculation and would seem irrelevant to the expressed goal of documenting improvement in these children of servicemen and servicewomen. The maturation explanation is not relevant for the reasons specified above, namely, that PDDBI T-scores are already age-standardized.

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

3. The hours argument is also illogical. Therapists are teaching children who lack social, play, self-care and language skills, who also often suffer from problems such as aversion to touch and/or sound, lack the ability to share attention with others, have difficulty generalizing what they have learned from one situation or person to the next, and who have no obvious motivation to acquire such skills often requiring the therapists to spend much time figuring out which rewards will effectively motivate the child to learn. I should note that the average number of hours required to learn a new language for a typical person without any of the above listed impairments ranges from 575 to 2200 classroom hours, according to the Department of State, depending on things such as language difficulty (https://www.rypeapp.com/blog/how-long-does-it-take-to-learn-mandarin/). How much more so is the case for a child learning to communicate his or her needs given the above-mentioned impairments?

In the summary found on pages 23-24 the following is stated:

"Overall, the findings from this analysis continue to demonstrate concern with overall outcomes of beneficiaries participating in the ACD. While the change scores in Figure 4 note improvements after 12 and 18 months of rendered ABA services, and that most baseline severity scores and most ages demonstrated some percent change in scores from baseline, the changes are small and may not be clinically significant. In addition, there is no comparison group (no treatment or another type of treatment) to note whether or not the change score at 12 months and 18 months is associated with ABA services or other treatments received. As a result, there is no way to know if the relatively small change observed here is the result of ABA services, other treatment, or if this simply is a result of maturation as noted in the PDDBI manual (page 60)....The reasons for these findings are not clear, but regardless of the reasons for these outcomes, ultimately, these findings demonstrate that the current format of the ACD, and the delivery of ABA services, is not working for most TRICARE beneficiaries in the ACD."

Comments:

1. The conclusions drawn here are inconsistent with the data as presented by the DoD in this report. The DoD demonstrates, in this problematic data set, rather large and dramatic changes in core signs of ASD as reported by parents in those cases likely to have ASD, especially in younger children and in those more severely affected. As presented, they appear to be clinically significant, as argued above, in terms of their likely impact on adaptive skills and overall severity of ASD but this could easily be determined by examining the effect sizes with appropriate statistical design and analysis of the extant data. They also underplay the actual "p values" obtained. The fact that this is not a controlled trial is obvious but, to this author, is besides the point as it is clear that many of these children are improving, and in direct contrast to the statement that the delivery of ABA services is not working for the TRICARE beneficiaries.

2. The PDDBI manual citation is a misreading of the section. **The changes described here in the domain and composite scores as a function of age reference the raw scores! The section cited by DoD provided the rationale for calculating the T-scores which take age of the child into account.**

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

The final conclusions with respect to the PDDBI can be found on p. 30 as follows:

> "Previous analyses yielded concerning results regarding treatment outcomes based on the scores reported in the Parent Form of the PDDBI. As a result, in preparation for this report, a more in-depth analysis of the available data was completed to include beneficiaries with baseline, 12-, and 18-month Parent Form PDDBI scores since the start of health care delivery with the West and East region MCSCs. The findings from this analysis continue to demonstrate concern with overall outcomes of beneficiaries participating in the ACD. While the change scores demonstrated small but statistically significant improvements after 12 and 18 months of rendered ABA services, and that most baseline severity scores and most ages demonstrated some percent change in scores from baseline, there was no comparison group (no treatment or another type or of treatment) to determine the attribution of these changes. It is also not clear if these changes are clinically significant. Subsequently, there is no way to know if the relatively small change observed here is the result of ABA services, other services received, or if this simply a result of maturation. However, the findings are clear that the number of hours of ABA services rendered did not improve symptom presentation of ASD based on the PAC scores. This finding strongly suggests that the small changes noted are not related to ABA services."

Comments and Conclusions:

1. What is evident in this summary is a minimization of the actual improvements reported in this document by referring to issues that don't appear, to this reviewer, relevant to the basic question as to whether families receiving services are seeing improvements in their children. It is clear that this is not a controlled trial and the lack of a "zero-order" correlation between reported hours of ABA services and outcome on core signs of ASD can be attributed to a multitude of factors including the obvious need to stratify this large sample not only by age and severity but by their interaction. In addition, there could be "lagged" effects over time, with changes in hours of service related to improvements at some later point in time. Overall, this is a complicated issue that cannot be answered as presented.

2. As presented, many children did improve in their core signs of ASD, many markedly so, despite the problems with the data and the data analyses so I do not understand why the DoD conclude that the changes are unrelated to ABA services. As they note, this is not a controlled trial. Such a statement would imply that, if there were a control group receiving the same ancillary services but without receiving ABA, then both groups would demonstrate the same level of change over time, other things being equal. The only thing that can be concluded is that many cases improved over time, especially in those cases where such improvement is particularly important (young children and very severely affected children).

3. As I noted above, it is also important to do the following in order to see which things changed in the children:
    a. Examine all of the domains and composite T-scores on the PDDBI.
    b. Match these outcome measures to the goals of intervention.
    c. Ensure that informants remain the same at all time points.
    d. Ensure that all PDDBI forms are completed and scored correctly as described in the

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

manual.

e. Develop appropriate statistical models that take relevant control factors into account such as age, sex, severity, family implementation, and goals of intervention into account; and

f. Examine effect sizes for these different outcome measures so as to understand their clinical meaningfulness

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Conclusions

As stated by the DoD, they are tasked with documenting the effectiveness of the ACD in children re-ceiving ABA services. They are doing this by monitoring PDDBI ratings of the children receiving such services over time. Their protocol does not involve carrying out a randomized experimental trial of ABA but, throughout the six reports reviewed here, they continue to make statements regarding the lack of efficacy of ABA even though they acknowledge that these conclusions are not based on such a controlled trial so there is an inherent contradiction here. Aside from this obvious contradiction, their conclusions cannot be justified due to the profoundly serious limitations in how the PDDBI was used in each of these six reports. **Most glaring is the total lack of understanding as to how the PDDBI is scored and interpreted and the obvious failure to read the manual.** This led to elimination of vast numbers of cases, **many of whom may have been responders** because of the incorrect assumption that a score of "0" mean that the item was "missing." Instead, it means that the behavior was not seen! In addition to this error which led to dropping of most of the original 14,000 cases, varying numbers of remaining cases across reports, and potentially biasing the data toward non-responders, there were other significant issues including: 1) failure to ensure uniformity in informants at each time point; 2) assuming that the Autism Composite was the most appropriate measure of progress without considering the goals of intervention and the corresponding need to examine the 10 other domains in the PDDBI in relation to those goals; 3) the lack of basic information about the demographic and clinical status of the cases and their families; 4) the total lack of information as to the statistics used to analyze the data that were examined and the apparent failure to use proper statistical analyses on the data; 5) the incorrect assumption that a change of 10 points in the Autism Composite T-score was necessary to demonstrate improvement; and 6) the failure to use the manual to understand the clini-cal meaningfulness of the various domains and composite scores on the PDDBI. **As noted above, the conclusions one draws from an analysis are only as good as the quality of the information that goes into the analysis. If the data are sloppy, so are the conclusions and their validity, a GIGO problem.**

Overall, these issues render the conclusions drawn by DoD uninterpretable. From a scientific and clinical view, this is a shame. **The 14,000+ cases receiving ABA services represent a potentially rich database of information that could be mined, in an epidemiological sense, to understand the various factors that are associated with significant change in the domains assessed by the PDDBI.** While not a randomized double-blind group study (which has its own problems which I will not go into here), the scientific data gleaned from the analyses would be a rich and valuable database for future controlled trials and would be an important contribution to the literature as well as to practitioners and the families they wish to help.

Technical Report to the National Coalition for Access to Autism Services (NCAAS):
Department of Defense Reports to the Congress Regarding the Autism Care Demon-stration (ACD) projects'

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

References

Cohen, I.L., Sudhalter, V., Pfadt, A., Jenkins, E.C., Brown, W.T. & Vietze, P.M. (1991a). Why are autism and the fragile X syndrome associated? Conceptual and methodological issues. **American Journal of Human Genetics**, 48(2), 195-202.

Cohen, I.L.,Vietze, P., Sudhalter, V., Jenkins, E.C. & Brown, W.T. (1991b). Effects of age and communication level on eye contact in fragile X males and non-fragile X autistic males. **American Journal of Medical Genetics**, 38, 498-502.

Cohen, I.L., Schmidt-Lackner, S., Romanczyk, R., and Sudhalter, V. (2003a). The PDD Behavior Inventory: A rating scale for assessing response to intervention in children with PDD. **Journal of Autism and Developmental Disorders**, 33(1), 31-45.

Cohen, I.L. (2003b). Criterion-related validity of the PDD Behavior Inventory. **Journal of Autism and Developmental Disorders**, 33(1), 47-53.

Cohen, I.L. and Sudhalter, V. (2005). **The PDD Behavior Inventory.** Lutz, FL: Psychological Assessment Resources, Inc.

Cohen, I.L. (2006). A neural network model of autism: Implications for theory and treatment. In D. Mareschal, S. Sirois, G. Westerman, and M.H. Johnson (eds). **Neuroconstructivism. Vol. 2: Perspectives and Prospects**. Oxford: Oxford University Pres1s, 227-260.

Cohen, I.L., Rovito-Gomez, T., Gonzalez, M., Lennon, E.M., Karmel, B.Z., and Gardner, J.M. (2010). Parent PDD Behavior Inventory profiles of young children classified according to Autism Diagnostic Observation Schedule-Generic and Autism Diagnostic Interview-Revised criteria. **Journal of Autism and Developmental Disorders**, 40, 246-254. Epub 2009, Sept 1.

Cohen, I. L., Liu, X., Hudson, M., Gillis, J., Cavalari, R. N., Romanczyk, R. G. et al. (2016). Using the PDD Behavior Inventory as a level 2 screener: a classification and regression trees analysis. **Journal of Autism and Developmental Disorders**, 46, 3006-3022.

Cohen, I.L. (2017). **PDDBI Professional Manual Supplement, Adolescent Normative Data.** Lutz, FL: Psychological Assessment Resources, Inc.

Cohen, I. L. & Flory, M. (2018). Autism Spectrum Disorder Decision Tree subgroups predict adaptive behavior and autism severity trajectories in children with ASD. **Journal of Autism and Developmental Disorders**. 49(4), 1423-1437.

Cohen, I.L., Liu, X., Lewis, M.E., Chudley, A., Forster-Gibson, C., Gonzalez, M., Jenkins, E.C., Brown, W.T., and Holden, J.J.A. (2011). Autism severity is associated with child and maternal MAOA genotypes. **Clinical Genetics**, 79, 355-362. Epub 2010, May 22.

Lane, B.R., Paynter, J. and Sharman, R. (2013). Parent and teacher ratings of adaptive and challenging behaviours in young children with autism spectrum disorders. **Research in Autism Spectrum Disorders, 7(10)**, 1196-1203.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. **Journal of Autism and Developmental**

Technical Report to the National Coalition for Access to Autism Services (NCAAS): Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)

Disorders, 30, 205–223.

Restrepo, B., et al., (2020) Developmental–behavioral profiles in children with autism spectrum disorder and co-occurring gastrointestinal symptoms. **Autism Research**. DOI: 10.1002/aur.2354

Sidman, M. (1960). **Tactics of Scientific Research: Evaluating Experimental Data in Psychology.** New York: Basic Books.

Technical Report to the National Coalition for Access to Autism Services (NCAAS):
Department of Defense Reports to the Congress Regarding the Autism Care Demonstration (ACD) projects'

Submitted 8/29/2020
Ira L. Cohen, Ph.D.

Use and Analyses of the PDD Behavior Inventory (PDDBI)