Article details: 2016-0137	
Title	Assessing Diagnostic Classification Accuracy of Neurobehavioral Disorder associated with Prenatal Alcohol Exposure (ND-PAE)
Authors	James L Sanders PhD, Rebecca E. Hudson Breen PhD, John Holland MD, Peter Koegler MD
Reviewer 1	Dr. Susan J. Astley PhD MS
Institution	Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA
General comments (author response in bold)	The manuscript is an important and timely contribution to the literature. It is well written and concise. My comments below are minor, but important to address. Throughout the manuscript, the authors describe the "gold standard" as "a multidisciplinary clinical assessment". More accurately, their gold standard is "the 2005 Canadian FASD multidisciplinary diagnostic guidelines". Throughout the manuscript, including the Abstract, this important distinction needs to be made. **We thank the reviewer for indicating this. The aforementioned "Gold Standard" of comparison is now described as the 2005 Canadian FASD multidisciplinary diagnostic guidelines, with recognition that no true gold standard for FASD diagnosis exists. p 2, 4, 7, 9** Below are specific recommendations/comments by Section. Title: Comparing Diagnostic Classification of Neurobehavioral Disorder associated with Prenatal Alcohol exposure (ND-PAE) with the 2005 Canadian Fetal Alcohol Spectrum Disorders Guidelines **We thank the reviewer for the recommendation. This is a more definitive title. p 1** Abstract Background: For accuracy, please restate as follows: The purpose of this study is to assess the classification of the DSM-5 criteria, using the 2005 Canadian FASD multidisciplinary diagnostic guidelines as the "gold standard" of comparison. **This has been revised as suggested. p 2** Abstract Methods: Please revise as follows: Eighty-two patients underwent multidisciplinary evaluations using the 2005 Canadian FASD Diagnostic Guidelines in Alberta between 2011 and 2015. Sixty were classified FASD. **This has been revised. p 2** Classifications from the Canadian FASD guidelines and DSM-5 criteria were moderately correlated (Cramer's V (82) = .44, p<.01). **This has been revised as suggested. P 2** The authors use the term "prenatal alcohol disorder" when no such term formally exists in the field of FASD. It would be more accurate to replace the phrase with "FASD". For example: While there is considerable overlap between the domains assessed in DSM-5 and those assessed in the Canadian FASD Guidelines, the DSM-5 criteria for ND-PAE were far less likely to identify individuals who met the Canadian neurobehavioral criteria for FAS, PFAS, and ARND. Although the neurobehavioural domains assessed by ND-PAE are supported in the research, there are limitations with its diagnostic structure that restrict the identification of individuals with FASD. **We thank the reviewer for pointing this out. Prenatal alcohol disorder has now been replaced with ND-PAE and FASD throughout the manuscript. All instances of prenatal alcohol disorder are removed. P 2, 4, 8, 12** Introduction: Revise as follows for accuracy: The purpose of this study is to assess the classification of the DSM-5 criteria, using the 2005 Canadian multidisciplinary FASD diagnostic guidelines as the "gold standard" of comparison. While there is considerable overlap between the domains assessed in DSM-5 and those assessed in the Canadian FASD Guidelines, it is hypothesized….. **This has been revised. p 4** Setting: Please revise as follows for accuracy: Significant deficits (2 or more standard deviations below the mean) in at least three domains is necessary for a diagnosis of FAS, partial FAS (PFAS), and Alcohol-Related Neurodevelopmental Disorder (ARND).[12] Confirmation of PAE is necessary for the PFAS and ARND diagnoses. **This has been revised. p 6** Design: The authors report "Two clinicians independently reviewed client files for evidence of meeting the

	diagnostic criteria for ND-PAE [18].” Please report which of the two clinician’s ND-PAE classifications are presented in Table 1. How did you classify ND-PAE when the two clinicians’ classifications were discordant? **Addressing discordant cases was specified, “Discordant cases (n = 8) represented a small portion of the sample and were discussed between the two clinicians and in order to reach agreement on classification.” P 8** Inter-rater Reliability: It is unclear what is meant by “90% agreement on DSM-5 criteria (Kappa = .79)? Which ‘criteria’? Do you mean the classification of a patient as ND-PAE or not ND-PAE was concordant across the two raters 90% of the time? Please clarify. **This has been clarified under the Inter-rater reliability heading, “Good inter-rater reliability was established (Kappa = .79) (23) based on 90% agreement between the two raters on ND-PAE classification and non-classification.” P 8** Classification Accuracy of ND-PAE: Revise as follows for clarity (e.g., replace the use of the term “clinical assessment” with reference to the Canadian guidelines). Dichotomous classifications using the2005 Canadian criteria (FASD, not FASD) and DSM-5 criteria (ND-PAE and not ND-PAE) were moderately correlated (Cramer’s V (82) = .44, p<.01). Against the 2005 Canadian Guidelines as the “gold standard”, a total classification accuracy of the DSM-5 was 61%. Of particular importance, all classification errors were false negative (n = 32), meaning that all errors were the result of non-classification in DSM-5 in the presence of classification using the 2005 Canadian Guidelines. **Reviewer recommendations were integrated into this section describing analysis, “Dichotomous classifications using the Canadian criteria (FASD or not FASD) and DSM-5 criteria (ND-PAE or not ND-PAE) were moderately correlated (Cramer’s V (82) = .44, p<.01). Against the 2005 multidisciplinary Canadian FASD Guidelines, the total classification accuracy of the DSM-5 was 61%. Of particular importance, all classification errors were false negative (n = 32), indicating that 32 cases obtained non-classification in DSM-5 in the presence of classification using the Canadian Guidelines. For this reason, the DSM-5 possessed inflated specificity (100%, 95% CI [87.7%, 100.0%]) and low sensitivity (47%, 95% CI [33.7%, 60.0%]).” P 8-9** Strengths and Limitations: Please revise the following paragraph as follows for clarity: The outcomes of this study are influenced by the use of the 2005 multidisciplinary Canadian FASD Guidelines as the “gold standard”. There are several guidelines in use around the world [10,11,15,16] without an unequivocal gold standard. The outcomes of this study would vary if one of the other guidelines had been selected to serve as the ‘gold standard’. In addition, the FASD classifications derived using the Canadian guidelines were derived prospectively by a multidisciplinary team. In contrast, the ND-PAE diagnoses were derived retroactively by two independent clinicians reviewing casefiles. The use of a comprehensive multidisciplinary assessment is important given that this approach considers additional factors such as external, developmental, or familial factors, independent of the criteria that may result in a diagnosis or non-diagnosis on a case by case basis. **We have revised the Strengths and Limitations section as follows “The findings of this study were impacted by three main aspects of the methodological design. First, it is important to bring to light that although the 2005 multidisciplinary Canadian FASD Guidelines were used as the reference standard, there are several guidelines in use around the world without an unequivocal gold standard. “** **We appreciate that the reviewer has pointed out the need to further expand on how the methodology of this study impacted the results. We have further elaborated on these points in the Strengths and Limitations section. P 10-12** The authors report: “In addition, the terms FASD and ND-PAE were used synonymously as disorders caused by prenatal alcohol exposure, given the lack of differentiation between the two terms in DSM-5 criteria”. The terms FASD and ND-PAE, for the purposes of this study, are synonymous. This is not a limitation. ND-PAE defines the functional CNS abnormalities that can result from prenatal alcohol exposure. Thus, ND-PAE IS synonymous with the functional CNS component of the Canadian FAS, PFAS and ARND diagnoses. **We thank the reviewer for this comment. We agree that the following sentence should not fall under the Strengths and Limitations section: “In addition, the terms FASD and ND-PAE were used synonymously as disorders caused by prenatal alcohol exposure, given the lack of differentiation between the two terms in DSM-5 criteria”. This has been deleted. P 12** I think it would help if the authors noted this in the Design section of the manuscript. This study does not address the growth, facial or structural CNS components of a diagnosis under the umbrella of FASD. It addresses only the functional CNS component of the ND-PAE, FAS, PFAS and ARND diagnoses under the umbrella of FASD. **We thank the reviewer for this recommendation. As a result we have expanded the analyses to demonstrate that there was no correlation between ND-PAE classification and pFAS/FAS, which include physical and facial features (see background and results). P 5, 9** The authors report “Accuracy of classification based on DSM-5 guidelines in this study is limited by lack of clarity of what constitutes impairment.” This is a true statement, but also an understatement. One of the most serious limitations of ND-PAE as defined in the DSM-5 is the absence of a case

	definition for what constitutes "impairment". Is "impairment" 1 SD, 1.5 SDs, 2.0 SDs below the mean? This failure to define "impairment" will result in highly variable and inaccurate ND-PAE diagnoses across clinicians. Introducing ND-PAE in the DSM-5 without defining "impairment" is essentially reverting back to the old gestalt approach to FASD diagnosis that the field of FASD abandoned back in 1997. **We thank the reviewer for bringing this to our attention. The sentence has been revised as follows: "Accuracy of classification based on DSM-5 guidelines in this study is limited by the absence of a case definition for what constitutes "impairment". This issue is further fleshed out in the Interpretation section.** **In addition, the emphasis of DSM on observable, descriptive symptomology has been described in the conclusion. P 11-12** To conduct the current study, the authors imposed the definition of impairment used by the 2005 Canadian FASD guidelines (2 or more SDs below the mean) on the ND-PAE diagnosis. But in so doing, they impacted the outcome of the study. The decision to impose the Canadian definition of impairment on the ND-PAE diagnosis is not a fatal design flaw, it just changes the nature of the study question. This study is not "assessing the classification accuracy of the DSM-5 criteria, using a multidisciplinary clinical assessment as the "gold standard" because the authors altered the DSM-5 criteria. They imposed a -2SD rule on the DSM-5 criteria. **We acknowledge that the results of our study were influenced by the use of -2 SD to categorize impairments in the DSM-5. The Strengths and Limitations section addresses how the cut-off level of -2 SD impacted the study. Where if, alternative cut-off levels (e.g. -1.5 SD, -1 SD) were applied to the DSM-5 criteria, the specificity and sensitivity analysis would yield different results. P 11** In so doing, the study question becomes, "What proportion of patients receiving a Canadian diagnosis of FAS, PFAS or ARND have CNS dysfunction with significant impairment (-2 SDs below the mean) in the domains required by the DSM-5 ND-PAE diagnosis (neurocognitive function, self-regulation, and adaptive function)? This is an interesting question and worthy of publication. But it would be inaccurate to imply the current study is "assessing the classification accuracy of the DSM-5 criteria, using a multidisciplinary clinical assessment as the "gold standard" of comparison." It's not. **We appreciate the reviewer suggesting that we revise our research question. This is a valid point in that our study compares the 2005 Canadian FASD guidelines to the DSM-5 after imposing a -2 SD. Thus, we agree that the use of the term classification accuracy is not fitting. Moreover, instances of "classification accuracy" have been avoided, where possible. The research question has now been revised. p 5** Imagine for a moment that you did not impose the -2SD criteria on the DSM-5 ND-PAE. If you asked two clinicians from the Canadian diagnostic team to retroactively review the records of these 82 patients and classify them as ND-PAE or not ND-PAE without any guidance for what constitutes "impairment", it is highly likely their classifications would have differed from those presented in Table 1. It is also highly likely that their inter-rater reliability would have been lower than 90%. But since your two clinicians were trained in the more rigorous Canadian FASD diagnostic system, it would have been near impossible for them to meaningfully classify the 82 patients as ND-PAE or not ND-PAE in the absence of -1, -1.5 or -2 SD criteria for what constitutes "impairment". On the other hand, I could imagine giving the 82 case files to 2 clinicians not trained in the Canadian FASD diagnostic system and ask them to classify the patients into two groups (ND-PAE and not ND-PAE) in accordance with the DSM-5 criteria. I would anticipate there would be a low inter-rater reliability and a higher proportion of the patients would receive a ND-PAE classification (since the -2SD rule was not imposed). If you then compared their ND-PAE classifications with the FASD classifications derived by the multidisciplinary team using the Canadian Guidelines, then and only then would you truly be "assessing the classification accuracy of the DSM-5 criteria, using a multidisciplinary clinical assessment as the "gold standard" of comparison." Table 1 would look very different. The sensitivity would likely have been a bit higher because a higher number of patients with FASD would have been classified as ND-PAE (because the bar for "impairment" would have been lowered). In addition, it is highly unlikely the specificity would have been 100%. Without a -2SD criteria imposed on ND-PAE, it is highly likely more patients without FASD would have been classified as ND-PAE. This in turn would have dropped the specificity to below 100%. How far below remains unknown. Although I am presenting this as an example to illustrate how this study was impacted by imposing a -2SD criteria on ND-PAE impairment, I am not advocating the authors revise their study. **It is true that if two independent reviewers classified patients as ND-PAE or not ND-PAE without any guidance of what constitutes impairment, results would look rather different.** **However, for the purpose of comparing between the two diagnostic systems, the application of a threshold level was conducive to strong inter-rater reliability between the two reviewers when assessing the case files, as well as generally accepted thresholds of deficits in neuropsychology.** I am strongly encouraging the authors to more thoroughly convey to the Readers how their study design impacted the outcomes. I am especially concerned that Readers will over-interpret the 100% specificity and think that is truly a reflection of the DSM-5 ND-PAE performance relative to a rigorous diagnostic system like the Canadian system. That would be incorrect. The 100% specificity is an artifact of the study design. **We appreciate the Reviewer's feedback. The Strengths and Limitations section now includes a further explanation of how our study design impacted the results. P 10-11** What would be keenly interesting and informative to Readers would be an additional table or text that reports the 82 patient's Canadian and DSM-5 diagnostic outcomes. In other words, what

	proportion of each Canadian FASD group (FAS, PFAS, ARND, and Other) received a diagnosis of ND-PAE. **We thank the reviewer for this suggestion. Table 1 has been revised with diagnostic outcome being further broken down into FAS, pFAS, ARND, and No-diagnosis. This point also prompted additional analyses. P 16** In addition, what proportion had an outcome of -2SDs on each of the ND-PAE domains? Here is a hypothetical example: The ND-PAE outcomes for each Canadian Diagnostic Classification are as follows: FAS (n = 1): ND-PAE (n=1, 100%), Neurocognitive Dysfunction at -2SDs (n = 1, 100%), Self-Regulation at -2SDs (n = 1, 100%), Adaptive Function at -2 SDs: Communication (n = 1, 100%), Social Communication (n = 1, 100%), Daily Living (n = 0, 0%), Motor Skills (n = 1, 100%). PFAS (n = 13): ND-PAE (n=8, 62%), Neurocognitive Dysfunction at -2SDs (n = 7, 54%), Self-Regulation at -2SDs (n = 8, 62%), Adaptive Function: Communication (n = 8, 62%), Social Communication (n = 2, 15%), Daily Living (n = 6, 46%), Motor Skills (n = 3, 23%). ARND (n = 46): ND-PAE (n=20, 44%), Neurocognitive Dysfunction at -2SDs (n = 10, 22%), Self-Regulation at -2SDs (n = 12, 26%), Adaptive Function: Communication (n = 15, 33%), Social Communication (n = 10, 22%), Daily Living (n = 22, 48%), Motor Skills (n = 24, 52%). Not FASD (n = 22): ND-PAE (n=0, 0%), Neurocognitive Dysfunction at -2SDs (n = 2, 9%), Self-Regulation at -2SDs (n = 5, 23%), Adaptive Function: Communication (n = 2, 9%), Social Communication (n = 0, 0%), Daily Living (n = 5, 23%), Motor Skills (n = 2, 9%) **In addition, Table 2 has been added to the manuscript. It depicts the proportion of patients with significant impairment (-2SD) in DSM-5 areas separated by Canadian FASD diagnostic classification (No-Diagnosis, FAS, PFAS, ARND). P 18** Revise Table 1 title and row and column headers as follows Title: Crosstabulation – 2005 Canadian FASD Diagnostic Guidelines and DSM-5 Column Header: FASD Diagnosis from Multidisciplinary 2005 Canadian Guidelines Row Header: ND-PAE Diagnosis from DSM-5 **We thank the reviewer for this recommendation. The title of Table 1 has been updated to "Diagnostic outcome comparisons between the DSM-5 and the 2005 Canadian Multidisciplinary FASD Guidelines". The headers have been revised as suggested. P 15**
Reviewer 2	Dr. Julie Kable PhD
Institution	Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA
General comments (author response in bold)	The authors have submitted a manuscript regarding an archival record review in order to evaluate the reliability and validity of the proposed symptoms included in the DSM5's conditions in need of further study. This type of study is vital to advancing the field's understanding of the disorder and is needed to move the disorder into the main body of the DSM5 where it would be recognized as a unique disorder. The impact of this will be helpful in documenting the public health care costs associated with prenatal alcohol exposure and assisting affected individuals and their families in obtaining appropriate mental health care. The manuscript provides very useful information but lacks some details needed before it is ready for publication. Lines 36 to 54 outline the proposed hypothesis but it is unclear why the authors anticipated low sensitivity from the literature that they provided. Please elaborate on what supports this prediction so that it does not appear their hypothesis was generated after seeing the results. **We thank the reviewer for bringing this to our attention. We have elaborated on our hypothesis by commenting on the schematics of both sets of criteria. P 4** The authors should provide a brief overview of the Canadian Guidelines used for diagnosis and the time period in which the individuals were diagnosed. The Canadian Guidelines have evolved over time and it is unclear which formulation was used. Given a third iteration is now available, future readers will have a hard time tracking what was done without this information. **A brief overview is now included in the Introduction. Individuals were assessed and diagnosed between 2011 and 2015. This detail is provided in the Methods section. Further, Canadian Guidelines has been replaced with "2005 Canadian Guidelines" for accuracy. P 3-4, 5, 7** Lines 21-23, the use of the word "domains" is somewhat confusing given that ND-PAE has "super-domains" that refer to multiple "domains" in this context. It might be better to refer to these as "areas" of functioning to avoid confusion to the reader. **The text has been revised to refer to "super-domains" when describing such.** The authors should provide basic demographic and physical characteristics of the sample so the sample can be appropriately characterized. **Basic demographics have been provided in the Patients' subsection (ie. Age, sex, location of assessment). P 6** Relative to Lines 19-32-The DSM5 does not make a clear distinction of "impairment" by design. How "impairment" is operationalized within a clinical and research context is left up to the individual applying the criteria within their specific context. This is true of all mental health diagnosis with the exception of intellectually disabilities. Conventions are then adopted by groups of researchers and clinicians based on instruments available to them, the cultural context within which the information is applied, the scientific information available to them regarding the validity of the convention adopted, and the resources available to provide care. This aids in implementing criteria across various cultural contexts. It is important to make this distinction as this will not change in any iteration of the criteria. It is also important to know that the criteria have to be able to be implemented by all mental health professionals (social workers, psychologists, psychiatrists, counselors, etc.) so traditional formulations used in describing FASD neurodevelopmental outcomes that are psychological test based are not able to be used in this context.

	We agree with the statement that impairment is operationalized within a clinical and research context might be left up to the individual applying the criteria within their specific context. Or that further, it might be guided by the FASD diagnostic criteria being applied. However, traditional FASD neurodevelopmental outcomes are embedded within the DSM criteria (i.e. memory, executive functioning, learning, and IQ). Several, though not all, of these criteria are typically assessed neurodevelopmentally. This issue is more clearly highlighted in the revision "there remains a disconnect between the approach to diagnosis in DSM-5 and ND-PAE critieria. Since the third edition of the manual, DSM has adopted an approach based on descriptive, observable symptomology (27,28). In contrast, FASD diagnostic approaches utilize neurodevelopmental data. While the domains assessed in ND-PAE and FASD diagnostic systems are largely harmonious, the impairment thresholds for both are different, with FASD systems using norm-based assessment, and ND-PAE on clinical judgment. However, several ND-PAE domains (i.e. learning, memory, IQ, executive functioning) are most appropriately assessed through neurodevelopmental assessment, while other domains (i.e. affect regulation) may be most appropriately assessed clinically." P 12-13 The term impairment is defined within the larger context of the DSM5 and refers to problems in living (ie. social, academic/vocational, etc. contexts). This should be acknowledged in the discussion section as well. The authors could point out the need for an agreed upon convention among practicing clinical providers (see Carmichael's recent publication on this) and empirical research to justify the convention but further elaboration of statements made in lines 11-15 is needed. **It is true that the "the term impairment is defined within the larger context of the DSM-5 and refers to problems in living (ie. social, academic/vocational, etc. contexts)". This point has been added to the Discussion. P 10** The authors' adoption of > 2 SD below the mean is one operationalization of these criteria and is perfectly fine within the context of this research question but is not justified by the IQ criteria specified in the DSM5 ND-PAE formulation as is stated in the manuscript and this should be made clear. **We have clarified our reasoning in applying > 2 SD below the mean for DSM-5 criteria no longer to be justified by the IQ criteria. It is commonly agreed upon in the field of neuropsychology that > 2D signifies a severe deficit whereas 1.5 SD or 1.0 SD may only indicate moderate dysfunction. The -2 SD cut-off level was chosen with the intention of classifying true cases of FASD while maintaining inter-rater reliability. P 7** Others have chosen different cut-off levels including 1.5 SD and 1.0 SD in evaluating the criteria (see presentations at last year's FASD Vancouver meeting and at the American Psychiatric Association meeting as I don't believe they have been published in full manuscript form yet but the abstracts are available for referencing). **This has been acknowledged in the statement, "However, others have used more moderate cut-off levels in translating norm-referenced measurement into descriptive clinical impairment for DSM-5 criteria" with reference to the FASD conference abstract. P 7** The authors should describe the tests available to them in each of the areas assessed and the number of participants who had information available. This can be put in a table format. It is unclear what they used to assess intelligence, EF, learning, etc. This is needed to clarify if the criteria are problematic or is the operationalization of the criteria inadequate. **We revised to provide the reader with the tests administered for each of the 9 CNS domains. Test names can now be located within the Methods section. P 5-6** When reporting the inter-rater reliability of 90%, the authors should clarify what the agreement is? Is that for making the overall diagnosis, endorsement of a given domain, or endorsement of a symptom? Clarity regarding this would be helpful and information regarding agreement of each would be helpful. The agreement rate should be discussed in light of the high base rate of the disorder in a sample of clinically affected children with a history of PAE. **This has been revised for clarity under the Inter-rater reliability heading, "Good inter-rater reliability was established (Kappa = .79) (23) based on 90% agreement between the two raters on ND-PAE classification and non-classification." P 8** In the previous work in this area, comparisons of agreement have been made using one adaptive functioning criterion rather than the 2 symptoms out of the 4 with one being either social or communication. Is it possible to provide this information from this project as well? The authors allude to the problem in their discussion but providing the data to the APA committees reviewing the reliability and validity information regarding the diagnosis would be more helpful. **This consideration has prompted additional analysis, which is described in the Background, Methods, and Discussion sections. Classification of adaptive functioning deficits between DSM-5 and the 2005 Canadian Guidelines were highly correlated (Cramer's V = .71). This is elaborated on in the revised discussion section, "Identification within the adaptive functioning super-domain in the DSM-5 was highly correlated to assessment of adaptive behaviour within the 2005 Canadian Guidelines. However, a high correlation between such domains would be anticipated. The DSM-5 adaptive functioning super-domain includes assessment of communication deficits and motor skills, which are considered separate domains in the 2005 Canadian Guidelines." P 10** In lines 16-21, the authors discuss potential age differences but provide no information regarding the differences that may have occurred within their own sample. Given that this is a rare dataset that has information on both adults and children, this would be helpful. What is the predictive validity relative to age group? **The following was added to the revision in the methods section, "Of those receiving an FASD related diagnosis (N= 60), 43 were children and 17 were adults. Chi-square tests assessing**

	frequency of diagnosis between children and adults were not statistically significant (p = .07) but trended toward a higher proportion of diagnoses for adults." P 6 The use of the term "prenatal alcohol disorder" is unclear/undefined. There is no such thing defined within the field and just creates ambiguity for the reader. **"Prenatal alcohol disorder" has now been replaced with "FASD" and "ND-PAE" throughout the document.**
Reviewer 3	Dr. Christine Lilley PhD
Institution	Sunny Hill Health Centre for Children, Vancouver, BC
General comments (author response in bold)	It's great to see data on this new category, which has been much talked about but little researched as yet. I appreciate the importance of the question, and feel that your data provides a good introduction to the question of how these criteria might impact the field. One comment on relevance: As you have noted, the DSM5 does not clearly specify how the symptoms described are to be measured, or how severe the impairments must be, so you made the decision to operationalize them as (a) clinical assessment including standardized testing and (b) scores more than 2 standard deviations from the mean. While this is a reasonable choice, and stated transparently, it's not obvious that everyone would make this choice, and I would like to see an explicit statement in the discussion the sensitivity and specificity are still unknown when assessment is based on interview and/or questionnaire data only, or when they are based on standardized testing with alternative cutoffs (-1 SD and -1.5 SD). If it was easy to calculate sensitivity and specificity with these alternative cutoffs in mind, the paper might have even more impact. The group that wrote the criteria has stated in conference presentations that they are in the process of running their own analyses on this point, and it would be interesting to see if a different group came up with a similar result. If this is difficult, you might simply state that others may wish to examine this as the sensitivity and specificity would be expected to be different in these other situations. **These comments have been integrated and clarified within the document. The statement "However, future research could further explore these impairment thresholds when applying different cut-off levels or when using interview/questionnaire data" was added to the strengths/limitations section. P 11** One minor thing: while clinicians often refer to a diagnostic 'hit,' I have been asked not to use this term by patient advocates who find it disrespectful, so you might consider editing the last clause of the introduction to simply read "the requirement for either communication deficit or impaired social communication and interaction." **This is a thoughtful comment and we thank the reviewer for bringing this to our attention. This has been revised to remove the term "hit" p 5**