Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: Genetic Testing for Developmental Disabilities, Intellectual Disability, and Autism Spectrum Disorder

Draft review available for public comment from December 16, 2014 to January 13, 2015.

Research Review Citation:.Sun F, Oristaglio J, Levy SE, Hakonarson H, Sullivan N, Fontanarosa J, Schoelles KM. Genetic Testing for Developmental Disabilities, Intellectual Disability, and Autism Spectrum Disorder. Technical Brief No. 23. (Prepared by the ECRI Institute–Penn Medicine Evidence-based Practice Center under Contract No. 290-2012-00011-I.) AHRQ Publication No. 15-EHC024-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2015. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2095
Published Online: June 30, 2015

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 1	General Comments	This is a comprehensive and well written Technical brief regarding genetic testing in patients with intellectual disabilities and autism.	We appreciate the reviewer's assessment.
KI Reviewer 2	General Comments	There is persistent and pervasive confusion between functional conditions/diagnoses, such as autism and intellectual disability which are developmental disabilities (DD) and etiologic conditions, such as fetal alcohol syndrome, Fragile X and Rett syndrome, that are considered risk factors, causative, associated with or etiologic for functional developmental disabilities. This creates a lot of confusion until p. 24 when the issue of phenotype versus genotype is brought up for the first time. It is a major deficit in permitting one to understand the document.	Based on the reviewer's comment, We revised the Background section to further clarify functional diagnosis (phenotype-oriented description of genetic disorders) and etiologic diagnosis (genotype-oriented approach to description of genetic disorders). The revision includes moving up part the content in p. 24 of the draft report that the reviewer had mentioned.
KI Reviewer 3	General Comments	I found the report to have significant flaws that compromise the conclusions to a significant degree. These are articulated in the specific sections below. I have also attached a copy of the report .pdf with comments embedded throughout the report not all of which are included in the comments below for the purposes of length.	Thank you. We have responded to each of the specific comments made by the reviewer in the appropriate sections of this document.
KI Reviewer 4	General Comments	Well-written, esp. considering the time constraints.	We appreciate the reviewer's comment. Given the very broad scope of work, the timeframe for preparing the Technical Brief is indeed constrained.
Peer Reviewer 1	General Comments	This overall well written brief contains some grammatical errors throughout, a copy editor should read the text before publication.	This report has been copyedited as the reviewer suggested.

Commentator & Affiliation	Section	Comment	Response
	General Comments	A summary of my main concerns/comments: 1) The use of CMA in diagnosing many different conditions resulting in ID is not stated clearly. This needs to be made very clear, so that insurance company reviewers understand the inherent value of this test.	The report includes the following sentence: "Medical genetics groups now recommend chromosomal microarray analysis (CMA) as a first-line genetic test to identify genetic mutations in children with multiple anomalies not specific to well-delineated syndromes, nonsyndromic developmental disability/intellectual disability (DD/ID), and Autism Spectrum Disorder (ASD)." We believe this sentence clearly states that CMA can be used to diagnose many different conditions resulting in ID.
	General Comments	2) Similarly, next gen sequencing is capable of diagnosing an almost unlimited number of conditions resulting in ID. It makes no sense to count how many times each condition is listed in GTR. Next gen sequencing can identify almost every single gene disorder.	Laboratories/manufacturers may use the next generation sequencing (NGS) technology to develop different tests, including whole genome sequencing (WGS), whole exome sequencing (WES), mitochondrial sequencing, and targeted sequencing (for different coding areas). Therefore, not every NGS-based test is capable of diagnosing an unlimited number of conditions resulting in ID. Developers bring NGS-based tests for different purposes to the market via the U.S. Food and Drug Administration (FDA) or Clinical Laboratory Improvement Amendments (CLIA) pathway. We believe it is informative to provide a count of these clinically available tests.
	General Comments	3) While I understand the focus on disorders mentioned by the key informants, this resulted in a random collection of conditions. Many other, often more common, conditions or groups of conditions were omitted.	As we note in the report, this report mainly focus on nonsyndromic DD/ID, and ASD. Some DD syndromes are included mainly because they are particularly of interest to stakeholders, not because they are more common than other syndromes. We believe we make the point clear in the report.

Commentator & Affiliation	Section	Comment	Response
	General Comments	Abstract: Under the common methods, array comparative hybridization and microarray are listed separately, these are essentially the same test. It may be helpful to add next gene sequencing methods as well, because this technology is rapidly taking over the field.	As noted in the report, we collected the test information from the Genetic Testing Registry (GTR). In GTR, "aCGH" and "microarray" are listed separately as methods used in genetic tests. We are not able to determine if those microarray tests include aCGH or only refer to single nucleotide polymorphism (SNP)-array tests. Therefore, we decided to report them separately, the way they were reported in GTR. Next-generation sequencing methods have been included in the "sequencing" category.
	General Comments	Later in the background, the term CMA is also used. Choose one term and stick with it as much as possible.	As we describe in the report, CMA includes aCGH and SNP arrays. We have checked the two terms used in the report to make sure an appropriate one is used in each context.
Peer Reviewer 2	General Comments	The authors of this technical brief present a comprehensive review of the available evidence to evaluate the clinical utility of genetic testing for Developmental Disabilities (DD). This work is a critical step in understanding the clinical utility given the prevalence of DDs and the recommendations to refer families receiving a DD diagnosis to pursue genetic testing.	We appreciate the reviewer's comment.
Peer Reviewer 3	General Comments	The authors tackle a very challenging problem of assessing the clinical utility and summarizing the clinical data on genetic tests. It is very challenging to assess a technical area in the midst of rapid change in technology, cost and knowledge.	We thank the reviewer for the comment.
Peer Reviewer 4	General Comments	Overall, I found this report to be a well-written and informative. Appendix B is very concise and clear. Appendix D clearly represents a lot of information (and work) in a useful format. I had a few minor comments, and a problem with a couple of points that I will attempt to clearly describe.	We thank the reviewer for the feedback and have responded to the reviewer's specific comments in the appropriate sections of this document.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 3	Abstract	Page vi. Markers are more typically used in the context of mapping studies. Would suggest the use of variants which is now commonly used to describe genetic changes from single nucleotide variants and small insertions/deletions (in/dels) to large copy number variants including variants large enough to be seen on standard karyotyping.	We changed the wording as the reviewer suggested.
KI Reviewer 1	Additional Questions	Quality of the Report: Superior	We appreciate the reviewer's assessment.
KI Reviewer 2	Additional Questions	Quality of the Report: Good	We appreciate the reviewer's assessment.
KI Reviewer 3	Additional Questions	Quality of the Report: Fair	We appreciate peer-reviewers' time and comments. We believe the revisions we have made based on reviewers' input make the report more solid and useful.
KI Reviewer 4	Additional Questions	Quality of the Report: Superior	We appreciate the reviewer's assessment.
Peer Reviewer 1	Additional Questions	Quality of the Report: Good	We appreciate the reviewer's assessment.
Peer Reviewer 2	Additional Questions	Quality of the Report: Good	We appreciate the reviewer's comment.
Peer Reviewer 3	Additional Questions	Quality of the Report: Fair	We appreciate the reviewer's time and comments. We are hoping the revisions we have done based on peer reviewers' feedback make the report more solid and useful.
Peer Reviewer 4	Additional Questions	Quality of the Report: Good	We appreciate the reviewer's assessment.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 1	Background	The background regarding advances in genetic testing was reviewed comprehensively. The discussion of which DD's were included was reasonably well written (pg1) However, children with "overt physical anomalies were excluded (line 46-47, pg 1) - What is concerning about this, is that the report goes on to include children with certain diagnoses (Angelman, Prader-Willi, Rubeinstein-Taybi, Smith-Magenis and Velocardiofacial syndrome who DO have physical anomalies BUT have an established diagnosis, which was probably made via genetic testing. Therefore the exclusion is for children who do NOT have a diagnosis, perhaps because they did not have a formal genetics evaluation or appropriate genetic testing. The logic here is somewhat fuzzy. It is also not clear that this also excludes children with internal anomalies which are not "overt"?	We have revised the text to clarify that our intent was to address use of genetic testing when a diagnosis is not apparent from the clinical presentation.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 2	Background	See above. There needs to be much more clarity on functional diagnoses versus etiologic diagnoses as this underlies much of the reasoning regarding testing and its utility to clinicians and families (which would appear to not necessarily coincide). It is appropriate to separate conditions in which there are specific features suggesting a diagnosis (phenotypic characteristics) in association with DD versus conditions where the findings are functional alone and offer no phenotypic clues. Testing in these two situations should be looked at separately.	We have revised the content regarding functional vs. etiologic diagnosis. We have also separated ID, GDD and ASD (as functional diagnoses) from the 8 DD syndromes included in the Technical Brief. The 8 syndromes are diagnosed based on phenotype (syndromic features) and genotype (association to certain genetic variants). However, separating testing situation by ID/GDD/ASD vs. the 8 syndromes is not necessary. As we state in the report, this Technical Brief primarily focuses on idiopathic, nondysmorphic, nonsyndromic ID, ASD and GDD. The 8 DD syndromes are included largely because, for patients with these syndromes, manifestations of GDD, ID, or ASD might be the main reason for the families to seek evaluation (when overt dysmorphic features have not been noted). So, the testing situation of interest for ID/GDD/ASD and the 8 syndromes is basically the same—etiologic diagnosis/differential diagnosis.
KI Reviewer 3	Background	The section is generally good in describing the conditions of interest. However, there are two significant weaknesses. The first is the omission that there are rare disorders where therapy can ameliorate or prevent the full manifestations of ID. While this may not meet a semantically precise definition of the word cure, the result is that the child can ultimately lead a normal productive life, so from the societal perspective this is a 'cure' and confers high clinical utility. Even though rare, to not mention this is a significant omission that presents a negative bias (see additional information in the next paragraph). Comments have been added to the .pdf to reflect where it would be appropriate to add this.	Suggestions made in the .pdf have been noted and edits made where appropriate. Details regarding these edits are outlined further in the Guiding Questions section of this document.

Commentator & Affiliation	Section	Comment	Response
	Background	The second weakness is the use of old literature estimates of the genetic burden in ID (most recent article 2010). The upper boundary is presented as 40% but that is now considered to be higher. A comment was included in the .pdf that provides additional information.	We strived to use the most recent data/references that we were able to identify in the report. We also checked the comment the reviewer made in the pdf file. However, the reviewer did not provide any reference to support the new numbers discussed in the comment.
	Background	On page 1 the paragraph starting at line 37 notes the 11 DD disorders were considered based on literature review and key informant interviews. None of the specific disorders selected has any treatments that reverse the intellectual disability. As noted above, there are genetic conditions for which tests are available where treatment can prevent or significantly reduce the development of intellectual disability. Several of these examples relate to interventions in conditions where the ID is associated with severe epileptic encephalopathy and include: Tuberous Sclerosis Complex presenting with infantile spasms where treatment with the anticonvulsant vigabatrin controls the seizures and results in significantly improved cognitive outcome; glucose transporter type 1 (GLUT-1) deficiency treated with ketogenic diet; pyridoxine deficiency treated with pyridoxine replacement and cerebral folate deficiency which responds to folinic acid. Here is a review that gives an overview of some of these conditions (Epilepsia. 2013 Nov;54 Suppl 8:45-50. doi: 10.1111/epi.12423. Therapeutic approach to epileptic encephalopathies. Vigevano F(1), Arzimanoglou A, Plouin P, Specchio N.) There are other examples that do not involve epilepsy response.	The scope of this report, including the 11 DD conditions, was determined based on the interest of key stakeholders. The DD condition discussed in the reviewer's example is beyond the scope of work. We agree with the reviewer that the findings of this Technical Brief only apply to the DD conditions included in the report. We have revised the Background section and the Summary and Implications section to clarify this point.
	Background	[Comment the reviewer made in the PDF document] Page 1. Excludes any genetic condition that causes intellectual disability but for which a treatment exists. This impacts the assessment of clinical utility.	We added a caveat regarding the applicability of the report's findings. In addition, as we clarify in the report, this Technical Brief is not a systematic review and does not perform any "assessment of clinical utility."

Source: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2095
Published Online: June 30, 2015

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 4	Background	Balanced, straight-forward, and concise	We appreciate the reviewer's comment.
Peer Reviewer 1	Background	Causes for DD not listed in the background: infection (maternal measles, rubella etc); maternal illness (untreated phenolketonuria).	We have added infection and maternal illness to the list of causes for DD.
	Background	It is clear that the stakeholders chose some specific conditions to be included, based on their personal interests. There is no other obvious reason to include these, but not others. One may want to combine some conditions under the term "microdeletion syndromes" (Williams, Smith-Magenis, velocardiofacial, and most cases of Prader-Willi and Angelman syndrome). Other important conditions, such as neurocutaneous disorders, structural brain anomalies, epileptic encephalopathies, rasopathies and metabolic conditions are omitted. These conditions can result in apparent ID without obvious malformation, they affect more patients than some of the rare conditions listed, and they need to be included.	As we note in the report, the main focus of this Technical Brief is nonsyndromic ASD, ID, and GDD. The 8 specific DD syndromes are included based on the interest of the key stakeholders. Other DD conditions mentioned by the reviewer are deemed to be beyond the scope of the Technical Brief. We assessed the reviewer's suggestion of combining some DD syndromes under the term "microdeletion syndromes," but decided not to make that move. We believe that most readers of this report (e.g., payers, clinicians) may find it easier to locate the information by defined clinical syndromes than by potential genetic etiologies of the disorders.
Peer Reviewer 2	Background	The background provides a clear, brief review of the importance, relevance, and rationale of this work. There are more recent genetic testing approaches that are not mentioned as such in the background (e.g. exome or targeted sequencing). Even though these are not being used clinically in the same vein as CMA, etc, these newer approaches are important components in genetic testing for DDs and warrants further mentioning. As mentioned by Key Informants whole exome and genome sequencing may be used with increasing frequency in the care of individuals with DD. Clarification for why these newer approaches are beyond the scope of work can be specified under the guiding questions section.	We mention sequencing in the Background section as an example of new genetic testing methods. Sequencing includes whole genome sequencing, whole exome sequencing, targeted sequencing and mitochondrial sequencing). They are certainly within the scope of this Technical Brief. Based on the reviewer's comment, we expanded the example and mentioned whole genome/exome sequencing in the Background section and the Guiding Question section. More detailed description of these genetic testing methods are provided in Appendix B of the report.
Peer Reviewer 3	Background	The background usefully addresses the need for evaluation of genetic testing in clinical practice in terms of clinical validity and implications for treatment.	We appreciate the reviewer's comment.

Source: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2095
Published Online: June 30, 2015

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Background	pp 1-2: The background information on the disorders included under developmental disabilities was concise and informative.	We thank the reviewer for the comment.
	Background	pp 3-4: The section on Genetic Testing for DDs addresses the increasing use of genetic tests to diagnose these disorders or determine a potential etiology, as well as some of the associated issues.	We thank the reviewer for the comment.
	Background	pp 4-5: Two minor points. The section on Availability of Genetic Tests and FDA/CLIA discussion feels like a finding to me – part of the description of the intervention (FDA and commercial status). I think this information would be useful at the top of page 12 re explaining your decision re LDTs.	We believe the discussion about the FDA/CLIA should stay in the Background section instead of being moved to the Findings section. The background discussion is intended to help readers understand the general context for genetic testing. We do not consider the content a finding for a report on genetic tests for developmental disorders.
	Background	The section on Evaluating Clinical Utility seems more related to Methods?	The section on Evaluating Clinical Utility sets a background for both Guiding Questions and Methods. Therefore, we determined to leave the content in the Background section.
KI Reviewer 1	Guiding Questions	The questions were thought out well and were comprehensive	We appreciate the reviewer's assessment.
KI Reviewer 2	Guiding Questions	Questions are appropriate, with the caveat above. Again, the questions should differ when there are phenotype clues versus functional developmental disabilities alone	We reevaluated each guiding question and deem them appropriate for DDs with or without phenotype clues.
KI Reviewer 3	Guiding Questions	[Comment the reviewer made in the PDF document] Page 2. Should be and. Chromosome aberrations are genetic.	We revised the wording as the reviewer suggested.

Commentator & Affiliation	Section	Comment	Response
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 2. With rare exceptions there are 'cures'. The newborn screening disorders are examples of this. There are also rare disorders such as the folate transporter disorders that can be cured with folinic acid. Alternatively, this could be stated as once ID is fully manifested there are no cures. I would prefer the statement with exceptions, given that the focus of the report is on clinical utility. The utility of prevention ID or reversing effects with treatment would have extraordinarily high utility give that resulting productive lifetime.	Based on this comment, we removed the sentence that had caused the concern.
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 2. This is probably true for autism.	The reviewer appeared to agree with us. No revision is made based on the comment.
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 3. This would likely now be judged an underestimate as the results of exome sequencing are finding 25-30% causal variants in children who have already had extensive diagnostic evaluations without a cause being identified (that is the 30-40% able to be diagnosed by other tests has already been excluded). The cited references are very old--2006 and 2010 which represents 1-2 generations of test improvements (not including exome sequencing)	The epidemiologic data cited in our report is from the most recent references that our searches could identify. Because the reviewer did not provide any source or reference providing more recent estimates, we are unable to update the numbers.
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 3. As noted above this bulleted list should include the rare curative interventions. Leaving these out perpetuates the myth that all of these conditions are incurable.	The scope of this report, including the 11 DD conditions, was determined based on the interest of key stakeholders. Because those conditions are beyond this Technical Brief's scope, we determined not to add those rare curative interventions in the bulleted list.
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 5. The term aberration is not in common use. Again would refer to variant. Could change to "...such as a gene or chromosomal variant."	We revised the phase as suggested.

Commentator & Affiliation	Section	Comment	Response
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 5. I don't understand this key question. Up to this point there has been no discussion of a so-called 'standard of care diagnostic strategy' other than standard karyotype which is a genetic test.	We revised the question to clarify standard-of-care diagnostic strategy.
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 7. Should explicitly state whether this includes whole exome and/or whole genome sequencing.	We made the revision as the reviewer suggested.
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 8. Health outcomes is used in two different senses in this and following paragraph which leads to some confusion. I would restate this sentence as Patient-centered medical outcomes and intermediate outcomes..."	We changed the phrase in the following paragraph from “health outcomes” to “outcomes” to avoid potential confusions.
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 8. Would restate this as "Relevant educational, social and behavioral outcomes may vary..."	We revised the sentence to “Relevant outcomes may vary…” to make the description more general.
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 8. To pull this together, I would restate this as, "These outcomes relevant to ID include..."	We changed the sentence to “Outcomes relevant to ID include…” based on a previous comment made by the reviewer.
	Guiding Questions	[Comment the reviewer made in the PDF document] Page 8. To clarify, I would conclude with this sentence. "For the purposes of this report, the medical, educational, social and behavioral outcomes will be collectively referred to as 'health outcomes'."	Adding this sentence becomes unnecessary after we revised the 2 paragraphs based on the reviewer’s previous 3 comments.

Commentator & Affiliation	Section	Comment	Response
	Guiding Questions	I find Key Question 1 confusing. On page 5 of the report Key Question 1 is stated as, “Does using a genetic test lead to improved health outcomes in patients with DDs compared to the standard-of-care diagnostic strategy?” Up to this point there has been no discussion of a so-called 'standard of care diagnostic strategy' other than standard karyotype which is a genetic test (see additional criticism in the comparator section below). Exclusion of standard karyotyping is curious given that it is a genetic test, has the largest body of evidence regarding the impact of the result (albeit not of the highest evidence standard) and has been in routine use for the evaluation of DD/ID for 50 years. If the reason for exclusion was that clinical utility for this has been demonstrated then a different question arises given that the primary benefit of CMA/aCGH over karyotype is an increased diagnostic yield (discussed in detail below). That question is, does the increased diagnostic yield identify disorders that contribute to clinical utility over and above standard karyotyping?	We revised the question to clarify standard-of-care diagnostic strategy. The scope of work was determined based on the interest of the stakeholders for this project. Standard karyotyping is not of interest to these stakeholders. We decided not to change Key Question 1 as the reviewer suggested. That key question potentially addresses multiple health outcomes. Diagnosis yield is only one of the outcomes of interest.
	Guiding Questions	In interventions on page 7, sequencing is mentioned but it’s not clear at this point whether whole exome or genome sequencing is considered although later in the paper some studies are considered and these are discussed in Appendix B. It would be worth clarifying here.	We made the revision as the reviewer suggested. Now whole exome or genome sequencing are explicitly mentioned.

Commentator & Affiliation	Section	Comment	Response
	Guiding Questions	On page 8, the section on comparators is quite weak. The two examples of comparisons are no genetic testing, which is a very appropriate comparator as it addresses a critical part of key question one, that is does genetic testing provide clinical utility when compared with no testing. The other comparators listed consist of standard educational and behavioral test batteries. These tests are designed for the purpose of charactering the type and degree of developmental and behavioral impairment. They are not purposed, nor do they have the ability to make an etiologic diagnosis, which is the purpose of a genetic test. The example of Angelman syndrome provided on pp. 14-15 is particularly apt, as there are at least 5 other genetic conditions (e.g. Pitt-Hopkins, Rett, Mowat-Wilson, Kleefstra syndromes) that have very similar developmental and behavioral profiles, however they are caused by different genetic abnormalities and have differences in associated medical problems meaning that condition specific guidance is different. These tests are important compliments to genetic testing, in that even in a well-characterized condition such as Down syndrome there is much variability of ability. These tests are administered in order to determine the individual child's developmental, educational and behavioral profile in order to develop an individualized program.	We have revised the Comparators section by incorporating the comments from several peer reviewers.

Commentator & Affiliation	Section	Comment	Response
	Guiding Questions	Finally, there is no mention of other non-genetic tests that are utilized in the diagnostic process. These include neuro-imaging, EEG and metabolic studies, although as pointed out by Moeschler and Shevell, clinical evaluation that includes family history, neurologic and dysmorphologic examination is also utilized. Setting the latter aside, several key articles present data on the diagnostic yield of neuroimaging and metabolic testing in the context of the evaluation of GDD. These include references #3 and 4, and an article by Curry et al. (Evaluation of Mental Retardation: Recommendations of a Consensus Conference Am J Med Genet 72:468-477, 1997) that fell outside the date range of the literature search but provides good data for the yield of these tests. This is a serious limitation, as neuroimaging and metabolic studies are frequently used in evaluation of children with GDD (without the scrutiny and oversight being applied to genetic testing) despite diagnostic yields that are significantly lower than CMA and no evidence of clinical utility that has been developed in a systematic way.	Nongenetic tests are beyond this Technical Brief's scope thus not mentioned in the report. However, we would include any study that compare a test of interest with a nongenetic test. If such studies exist, our search would have captured it.
	Guiding Questions	The Outcomes section (p. 8) uses the term health outcomes in a confusing way. I have provided suggested language in the .pdf.	We changed the sentence to "Outcomes relevant to ID include…" based on a previous comment made by the reviewer.
KI Reviewer 4	Guiding Questions	Outline structure works, but numbers from diagram could be more clearly connected (bolding or other)	We evaluated the reviewer's suggestion and feel that the numbers in the diagram seem to be well marked. Bolding the numbers may not further improve the diagram's clarity. So we elect not to further edit the diagram.
Peer Reviewer 1	Guiding Questions	Under the "Interventions": CMA (and aCGH) identify copy number variations, NOT MUTATIONS. The same is true for SNP array, but this test may in addition identify regions of homozygosity.	We changed the wording from "genetic mutations" to "genetic variations."

Commentator & Affiliation	Section	Comment	Response
	Guiding Questions	It is very important to outline what CMA does: It can diagnose any microdeletion syndrome. Because this technical brief will likely be reviewed by insurance companies to make coverage decisions, the clinical use of CMA needs to be explained. This is not a single test ordered for a single diagnosis. It is a test capable of identifying a large number of conditions resulting in ID. Its importance cannot be overstated.	As we describe in the Background section and Appendix B, CMA may detect microdeletions, microduplications, and other copy number variants. We are aware that CMA, as a "testing method," is capable of identifying a large number of genetic variations potentially resulting in IDs. Many CMA tests have been developed targeting different genes or chromosomal regions for different clinical conditions. These tests are not considered one test, although they are based on the same genetic testing method.
	Guiding Questions	Again missing from the list of tests under review is next generation sequencing. This is so important that it needs to be included.	We added whole genome sequencing and whole exome sequencing to Guiding Question 1b. These sequencing tests may use the Sanger method or next-generation sequencing platforms.
	Guiding Questions	The comparators are a very difficult concept. While I can follow the logic used for listing those for ASD, the comparators for Angelman syndrome do not make sense. A clinical diagnosis of Angelman syndrome may include a developmental test, a physical examination by a dysmorphologist/geneticist and an EEG to identify the characteristic EEG pattern. For many of the genetic conditions resulting in ID, there are simply no meaningful comparators to the appropriate genetic studies.	We appreciate the comment and re-wrote the *comparators* sections to reflect the broad range of possible comparators given the range of conditions.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Guiding Questions	The authors provide clearly operationalized guiding questions for the work. However, in the background section, the authors list 7 key questions under the framework for establishing the chain of evidence for clinical utility. It would be helpful if how these guiding questions fell under the above framework was described or specifically how these guiding questions fed into the responses to the "framework" questions. As mentioned above there is It would be helpful if how these guiding questions fell under the above framework was described or specifically how these guiding questions fed into the responses to the "framework" questions. .	We removed the evaluation framework and the associated 7 key questions to avoid any confusion with the guiding questions.
Peer Reviewer 3	Guiding Questions	1. Comparators are dx of ID or ASD, not of association with genetic condition. Using comparators to make a phenotype diagnosis isn't relevant. For examples of chromosomal disorder, the comparator would be Conventional G-banded karyotyping.	Based on the comments from several reviewers, we revised the Comparators section.
	Guiding Questions	2. Sequencing should be clarified in abstract as directed Sanger sequencing and WES or CGS should be stated as not in the scope of the review, based on the review's scope as currently written.	We revised the guiding question to clarify the sequencing tests relevant to this report.
Peer Reviewer 4	Guiding Questions	The Guiding Questions are clear and the PICOs are helpful in specifying the scope. There was no indication of changes. I did not see an answer to Guiding Question 2b (apologies again if I missed it). Of course there was likely no source since almost entirely LDTs, but should note. Same for 2c – again, such information is not likely to available, but useful to complete.	In the tables in Appendix D, we reported information on "Genetic Counseling Required Pre-Test or Post-Test" that we collected from the GTR database. We did not identify other information addressing Guiding Questions 2b and 2c. Based on the reviewer's comment, we revised the report to describe this finding more explicitly.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 1	Methods	A detailed review was undertaken and determined testing which is available across the country. Online sources were used as well as "gray" literature sources including specialty society Web sites and clinical tiral databases. A review was undertaken of published literature, but the entire article was not generally reviewed, only the abstracts were read. (pg 16-17) Articles were reviewed which have been published since 2000 (pg vi). Genetics is such a rapidly changing field that for many papers and reviews, it is recommended that one should look at what has been published more recently (ie within the past 5 years).What might have been published as late as 2008 or 2009 may no longer be relevant in 2014-15.	As we describe in the Methods section, we collected a portion of the data for this report by reviewing abstracts. But in cases in which abstracts provided insufficient information, or in which there was reasonable uncertainty regarding the required data, we retrieved and reviewed full-text articles. We agree with the reviewer that genetics is a rapidly changing field and we should focus on more recent literature. However, determining an appropriate cut-off date for the literature search was not straightforward and required subjective judgment. After consulting with the key informants for the report, we decided to search for studies published since 2000.
KI Reviewer 2	Methods	Methods appear largely ok, again noting the mixing of "apples and oranges". Also it is sad but true that the research done in other countries on this topic is more robust and perhaps should not have been dismissed so readily.	We revised the content regarding functional vs. etiologic diagnosis. The revision should have addressed the reviewer's "apples and oranges" comment. Although our search was limited to English language studies, this Technical Brief includes studies conducted in other countries except for economic analysis. In addition to the difference in currencies, many other factors may affect the transferability of economic evaluations across countries. There is a rich body of literature addressing this issue (e.g., Goeree et al. 2007 [PMID: 17407623]; Drummond et al. 2009 [PMID: 19900249]). Because this Technical Brief is primarily to inform U.S. stakeholders, we believe it is appropriate to exclude foreign cost-effectiveness analyses.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 3	Methods	The description of data sources is good. There are two other sources that were not used that collect more complete information on clinical validity and in the case of the latter also information on clinical utility. These are the New York State Genetic Testing registry and the CMS MolDx program. The latter is more recent so depending on when the data collection ended, this may not have been available. The former has been registering genetic tests for over 10 years and most national and specialty genetic testing laboratories have New York State certification because this is needed for any test done on a NY state resident. These sources, (particularly NY state) do not have the limited information that led to the exclusion of the other sources. This omission of these gray data sources weakens the methods and should be reflected as a potential weakness.	Dr. Michele Caggana, the head of the Genetic Testing Section for the New York State Department's Clinical Laboratory Evaluation Program, is one of the Key Informants we consulted for this Technical Brief. We double checked with her about any "New York State Genetic Testing registry." Dr. Caggana was not aware of such a registry. She also agreed with our gray literature search strategy (i.e., using the NCBI's Genetic Testing Registry as the primary source to identify tests relevant to this Technical Brief). The MolDX program managed by Palmetto GBA is one of the gray literature sources that we explored in the early phase of this Technical Brief project. This program primarily focuses on reimbursement coverage policies related to genetic testing. The program's Web site does not provide detailed information about genetic testing that we are looking for.
	Methods	The beginning data of the literature search is provided (2000) but the end date isn't. This should be added. I would give year and month, as there are a couple of case series published in the last half of 2014 that look at changes in patient management as a result of whole exome sequencing (Soden et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. Science and Translational Medicine 6:265ra168; Srivastava et al. Clinical Whole Exome Sequencing in Child Neurology Practice Annals of Neurology 76(4):473-83)	We have updated our literature search to February 2015. Both the Soden and Srivastava studies the reviewer mentioned have already been added to the report. We added the end date for the literature search period as the reviewer suggested.

Commentator & Affiliation	Section	Comment	Response
	Methods	Literature search: In the abstract (p. vi) the term 'systematic search' was used. In this section it is noted that "…a complete review of all full-text articles was not feasible. We therefore collected a portion of the data for this report from a review of abstracts." Given this I think that while the term systematic search is technically accurate, it is somewhat misleading, as the reader may assume this was a systematic review rather than search. To reduce this confusion I would suggest using the term structured literature search. This should also be changed in the limitation section.	As the reviewer commented, what we did is a systematic literature search. We think it is appropriate to identify our search effort as "systematic search." In various sections of the document, we clarify that this report is a "Technical Brief," not a "systematic review." We also discuss the limitation of the report as a technical brief and a potential need for a systematic review in the future. We believe the chance for readers to mistake the report for a systematic review is minimal. Based on this comment, we revised the abstract to clarify the point from the beginning of the report.
	Methods	While I understand that the Regier paper was excluded as a foreign study, I think this is a mistake. This is the only study that models the impact of the increased diagnostic yield of array CGH compared to conventional cytogenic testing. While the unit of measure was the Canadian dollar, the transitional probabilities used were taken from the published literature (with some supplement from local data), therefore would be the same irrespective of the country from which the study was done (with the exception of countries with a high rate of consanguinity). While increased diagnostic yield is not equivalent to clinical utility, many of the health outcomes studied in this report depend on an accurate diagnosis, therefore are not separable from the diagnostic rate. [Regier DA, Friedman JM, Marra CA. Value for money? Array genomic hybridization for diagnostic testing for genetic causes of intellectual disability. Am J Hum Genet. 2010 May 14;86(5):765-72. PMID: 20398885]	In addition to the difference in currencies, many other factors may affect the transferability of economic evaluations across countries. There is a rich body of literature addressing this issue (e.g., Goeree et al. 2007 [PMID: 17407623]; Drummond et al. 2009 [PMID: 19900249]). Because this Technical Brief is primarily to inform U.S. stakeholders, we believe it is appropriate to exclude foreign cost-effectiveness analyses.
	Methods	[Comment the reviewer made in the PDF document] Page 10. Should provide the end date for the literature search. This was not listed in the AHRQ description of the technical brief provided and is crucial for the reader to determine how current the results are, given the rapidly changing nature of the evidence base.	Because the searches were updated during the peer review phase of the project, we have now added the end date for the literature search period.

Source: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2095
Published Online: June 30, 2015

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 4	Methods	Reasonable (if taken with appendices)	We thank the reviewer for the comment.
Peer Reviewer 1	Methods	The described methods are reasonable, given the scope of this brief.	We appreciate the reviewer's assessment.
Peer Reviewer 2	Methods	The methods are the appropriate methods to address the stated guiding questions. Similar to linking the guiding questions to the evidence framework questions in the background, it would be helpful to link particular methodological approaches to the specific guiding questions.	Based on the reviewer's comment, we revised the report to link the methods to guiding questions. The methods described in the section (including seeking input from Key Informants, searching gray and peer-reviewed literature, literature screening) generally apply to all guiding questions. However, we relied more on gray literature search to address Guiding Questions 1 and 2, and relied more on peer-reviewed literature addressed Guiding questions 3 and 4.
Peer Reviewer 3	Methods	The authors are only taking voluntary submission to The GTR rather than expecting peer review of validation of assays. They note this as a limitation but the table needs to highlight this problem or someone may think this information has clinical utility and make errors.	In the Methods section, we explicitly describe that GTR data were submitted by test developers on a voluntary basis. Later in the report, we also discuss using data voluntarily submitted to GTR as a major limitation of this Technical Brief. We further put footnotes under Tables 1, 2 and 3 to clarify the data are from GRT. In addition, we mention the following in the Findings section: "Our search of the GRT database identified a limited amount of data on analytic validity or clinical validity for a portion of the 672 tests. However, references were rarely provided for determining where these data came from. We deemed these data to be unreliable and did not to report them in this Technical Brief." We believe there is a minimal chance for readers to conclude that the summarized GTR data address the clinical utility of the tests

Commentator & Affiliation	Section	Comment	Response
	Methods	The authors should mention the role of ISCA and its use as a source for interpretation of array results and the review would be improved by at least demonstrating its use in a figure.	In the Summary and Implications section of this report, ISCA [the International Standards for Cytogenomic Arrays] is listed as one of the public databases that can be used to facilitate the identification of causal genetic aberrations. We deem it unnecessary to demonstrate use of ISCA in a figure.
Peer Reviewer 4	Methods	1) I think the section on the role of KIs could be more informative. Was there general agreement in the input received on the current role of these tests in clinical practice and the potential advantages or harms, or was the feedback diverse and/or polarized based on varying perspectives? You noted that KI input was identified throughout, but I found it mentioned only one other time in the subsequent paragraph (apologies if I missed it).	As we describe in the report, the role of KIs is to serve as a resource to offer insight and help us identify important issues from different perspectives. The goal of the discussions with KIs is not to reach any consensus. We assessed different views from the KIs and incorporated them into the report as deemed appropriate. Our discussions with the KIs were documented in the KI call summaries and have been submitted to AHRQ.
	Methods	2) Gray Lit and Lit review and Data Abstraction sections were concise/clear.	We appreciate the reviewer's comment.
	Methods	3) I do have a concern about the published literature search with regard to identifying articles addressing analytic and clinical validity and clinical utility. Reviewing Appendix A, I noted that no search appeared to include relevant terms such as 'analytic validity', 'analytic sensitivity/specificity', 'clinical validity/sensitivity/specificity', 'detection rate', 'false positive rate', 'clinical effectiveness', or 'clinical utility'. I admit I have no idea if it would have made a difference, but it makes me wonder if the number of articles might have been underestimated. The terms 'sensitivity', 'specificity', 'ppv' etc. did certainly work to some extent as you found 21 papers.	The relevant terms for literature search that the reviewer brought up were actually included in our original search strategy. Please refer to set #19 in the Embase/Medline search strategy, set #17 in the PubMed strategy, and set #29 in the Cochrane strategy in Appendix A of the report.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 1	Findings	1. The types of tests available were described in great detail, and are confusing, especially for the non-geneticist. For example, 113 tests were identified to diagnoses "Angelman syndrome". In actuality, there are 4 main methods used to diagnose this disorder: using FISH or microarray to detect a deletion of chromosome 15, methylation studies to rule out imprinting errors, and actual analysis of the UBE3A gene. When one looks for example at Univ of Chicago, pg D-43, multiple other genes are listed (lines 28-44), these are genes which cause other disorders which may have some manifestations which look like Angelman syndrome, but are in fact other diagnoses. See also my comment #4 below regarding how one decides on which tests to do first. The same problem is seen for Fragile X syndrome, which lists 56 different tests identified for making this diagnosis, when actually, it is one test - analysis of the FMR1 gene (which has several different changes to it, but NOT 56 different abnormalities causing this syndrome)	The two tests offered by the University of Chicago that the reviewer mentioned analyze multiple genes including UBE3A. As the reviewer pointed out, analysis of UBE3A is one of the testing methods for evaluating Angelman syndrome. We think it is appropriate to include the two tests in the report. We believe it is appropriate to identify the 56 fragile X syndrome-related tests as 56 different "laboratory-developed tests." Although these tests analyze the same gene—FMR1, they are developed by different laboratories using potentially different testing methods, including next-generation sequencing, Sanger sequencing, PCR, microarray, Southern blot, and aCGH. So, these 56 tests are not one same test. Similarly, we believe it is appropriate to identify 113 different Angelman syndrome-related tests as 113 different tests.

Commentator & Affiliation	Section	Comment	Response
	Findings	2. The study found that insufficient evidence for clinical benefit of genetic testing has been reported. What disturbs me about this finding, is that the methodology of this study would make it more difficult to find this evidence for selected, specific disorders. It feels like all testing has been grouped together in aggregate, and that might lead to the interpretation that no testing for any underlying diagnosis is worthwhile. This would be problematic. Of all the articles listed in Table 4 Clinical Utility studies, there is only one (Ellison 2012 line 11-16) which actually looks at changes in clinical management. There are in fact MANY changes in clinical management which would occur if a confirmed diagnosis was made in Angelman, Smith-magenis, VCFS and Williams syndrome, as well as a number of other diagnosis which were not covered) For example, a child diagnosed with a PTEN mutation would have very clear changes in medical management, changes in medical management for a carrier parent, and clear genetic recurrence risks which would be critical for the family.	The report certainly does not conclude or suggest that no testing in this area is worthwhile. Actually, in various sections (Background, Summary and Implications), the report discusses potential benefits of genetic testing for assessing DDs. We also disagree with the comment that we have grouped all testing in aggregate. As Tables 1-6 in the Findings section and the tables in Appendix D clearly show, the report provides sufficient information for differentiating the findings on the tests and evidence identified by DD conditions, testing methods, and targeted genetic variations. Table 4 includes not only one but six studies (Srivastava 2014, Ellison 2012, Iglesias 2014, Mroch 2012, Coulter 2011, Bruno 2009) that assessed the impact of genetic testing on clinical management of patients with a DD condition. The table summarizes peer-reviewed published evidence addressing clinical utility that we identified using the structured search strategy (see the Methods section of the report). The reviewer discussed other potential changes in clinical management based on genetic testing but did not provide any references. Therefore, we are not able to add additional studies to the table based on the comment.

Commentator & Affiliation	Section	Comment	Response
	Findings	3. I worry that the importance of being able to provide accurate recurrence risks to a family has not been emphasized. this is critical for family planning.	We agree that accurate recurrence risks is an important piece of information for family planning. One goal of our literature search and screening was to capture studies that addressed the impact of genetic tests (including those for predicting DD recurrence risk) on family decisions (including decisions related to family planning). In fact, two studies in this category (Iglesias 2014; Bruno 2009) were captured by our search and are summarized in Table 4. However, we are unable to further emphasize the issue beyond what our literature search has identified.
	Findings	4. I don't think that the different modalities for confirming a diagnosis were explained in such a way that it would be clear to a non-geneticist - For example, there are many ways to confirm a diagnosis of Angelman syndrome, but it is NOT clear that this should be done in a stepwise fashion and what that would look like.	This report includes a broad range of development disorders, and clinicians/geneticists do not always agree on the diagnostic approaches for these disorders. Instead of offering too much discussion of the diagnostic approaches that reflects our own viewpoints, we summarized consensus statements or recommendations (Table 7) from clinical practice guidelines regarding how to use genetic tests in DD diagnosis. We are hoping these guidelines provide more balanced views.
	Findings	5. The GTR database which was used cannot distinguish between older and newer versions of the test, or explain why a lab might use multiple tests for the same dx (pg 12, line 50-52)	We removed the sentence that caused the confusion.
KI Reviewer 2	Findings	Findings are no surprise given the known limitations of available research. The definitions and explanations of the various types of tests are well done.	We appreciate the reviewer's assessment.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 3	Findings	Proposed Intervention: On page 18, the paragraph beginning on line 38 discusses issues around diagnostic yield. The points made are reasonable, but appear to overemphasize the potential disadvantages of using this as an intermediate outcome. Since the introduction of karyotypes in the early 1960s, genetics has been faced with rapid increase in diagnostic technologies, each of which yields additional diagnoses with the attendant difficulties noted by the authors. A prime tenet of genetics is that the wrong diagnosis is worse than no diagnosis, so putative etiologic diagnoses are scrupulously evaluated with a conservative approach to definitive assignment. This hasn't been quantified in the literature, so the critique is not without some merit. However, since condition specific management plans are being considered as an intermediate outcome that bears some relation to utility and given that condition specific management cannot be initiated without a specific diagnosis, improved diagnostic yield is important and related to improvements in patient management.	We do not agree that we have overemphasized the disadvantages of diagnostic yield studies. We believe it is reasonable and important to discuss those issues related to interpreting the findings of diagnostic yield studies. We agree with the reviewer that improved diagnostic yield may have impact on patient management. We revised the Summary and Implications section to reflect this point.
	Findings	Evidence map: I found no discussion relevant to Key question 1 that is genetic diagnostic testing compared to the standard-of-care diagnostic strategy or no testing. As noted above, the chosen comparators are not appropriate in my view, but there is no discussion of the evidence for genetic testing outcomes compared to the chosen comparators. This seems a major omission.	We disagree with this comment. Searching for evidence that addresses Key Question 1 is actually a main focus of this Technical Brief. In the Findings section, we report that we did not identify any studies that directly evaluated the impacts of a genetic test on health outcomes (i.e., addressing Key Question 1). In the Summary and Implications section, we further discuss this finding and describe it as a main evidence gap for addressing genetic testing's clinical utility. Per comparators for Key Question 1, we revised the question to clarify standard-of-care diagnostic strategy.

Commentator & Affiliation	Section	Comment	Response
	Findings	A comment about an omission in the discussion of clinical validity. In the section where CV is discussed and in Table 9 (Evidence Gap) it is noted that there are no comparative studies of tests "…that evaluate the test diagnostic accuracy using a gold standard or other acceptable reference methods" The unacknowledged challenge is that all new genetic diagnostics detect things that were previously undetectable, therefore there is no gold standard or other acceptable reference. This should be stated explicitly lest the reader conclude that such tests exist and the field has not done due diligence in not performing these studies.	We revised Table 9 regarding ideal evidence for addressing clinical validity. The wordings for ideal evidence have been changed to "cohort studies that evaluate the test's diagnostic accuracy using phenotype or other acceptable diagnostic standards as reference methods." After the revision, the reviewer's comment on Table 9 no longer applies.
	Findings	[Comment the reviewer made in the PDF document] Page 12. Should be the American College of Medical Genetics and Genomics. Also, the ACMG-CAP program is a joint program, so and would be appropriate rather than or.	We thank the reviewer for identifying the inaccurate content. We have revised the content as the reviewer suggested.
	Findings	[Comment the reviewer made in the PDF document] Page 13. As noted previously, the MolDX and New York State Certification program has reference information and some independent assessment and validation of laboratory submissions.	Dr. Michele Caggana, the head of the Genetic Testing Section for the New York State Department's Clinical Laboratory Evaluation Program, is one of the Key Informants we consulted for this Technical Brief. We double checked with her about the possibility of using any data from New York State for this Technical Brief. Dr. Caggana did not think any data from her program would suit the need of this Technical Brief. She agreed with our gray literature search strategy (i.e., using the NCBI's Genetic Testing Registry as the primary source to identify tests relevant to this Technical Brief). The MolDX program managed by Palmetto GBA is one of the gray literature sources that we explored in the early phase of this Technical Brief project. This program primarily focuses on reimbursement coverage policies related to genetic testing. The program's Web site does not provide detailed information about genetic testing that we are looking for.

Commentator & Affiliation	Section	Comment	Response
	Findings	[Comment the reviewer made in the PDF document] Page 17. This may be a transcription error. It should be MECP2.	Yes, this is a typographical error. We have corrected it.
	Findings	[Comment the reviewer made in the PDF document] Page 27. As noted in the methods critique, while the comment about the economic and health care systems being different is correct, the information related to evidence about the utility of the tests is still of some relevance.	This Technical Brief is primarily to inform U.S. stakeholders, thus we believe it is appropriate to exclude foreign cost-effectiveness analyses. Many factors may affect the transferability of economic evaluations across countries. There is a rich body of literature addressing this issue (e.g., Goeree et al. 2007 [PMID: 17407623]; Drummond et al. 2009 [PMID: 19900249]). Clinical utility of genetic tests is not the subject of the economics-related section on p.27. If the foreign studies in question provide relevant data on clinical utility, they would be captured by our search and be summarized in Table 4.
KI Reviewer 4	Findings	"reported outcomes" in table 4 could be expanded or maybe done with bullet points/outlines.	Only 1 study (Makela 2009) in Table 4 reported more than 1 outcome. We separated the outcomes for that study as the reviewer suggested.
Peer Reviewer 1	Findings	"Some laboratories offer multiple tests for the same DDs (e.g., ID, ASD). These tests differ in gene markers targeted, analysis methods used (e.g., sequencing, microarray), or the purposes of testing (e.g., screening, diagnosis). The information provided by the GTR database is not sufficient for judging whether any of these tests is the newer version of another test or why these laboratories offer multiple tests for the same DDs" This paragraph illustrates a lack of understanding regarding the numerous different causes for conditions presenting with ID, as well as a lack of understanding of the appropriate order of studies (such as CMA; subsequently sequencing of single genes OR methylation study for Prader-Willi syndrome; possibly followed by a next generation sequencing approach).	We were aware that a DD may have multiple genetic causes and that a laboratory may develop separate tests targeting these different genetic causes. Based on this comment, we revised the paragraph and removed the sentence that caused the confusion.

Commentator & Affiliation	Section	Comment	Response
	Findings	Table 2 illustrates a number of problems with the way tests are counted. For example, every single CGH or microarray will diagnose Williams syndrome and velocardiofacial syndrome, both are microdeletion syndromes that ALWAYS are identified on these tests. Why do they not show the same number of tests? This is only due to the lab bothering to list this as a single indication. It is VERY important to make it clear that a SINGLE test (microarray) can identify NUMEROUS genetic conditions. This message is nowhere to be found in this brief.	In this report, aCGH/microarray is considered a testing method, not a clinical test. A clinical test is either an FDA-cleared/approved commercial kit or a laboratory-developed test (LDT) that is clinically available in the United States. Labs/manufacturers may develop different tests using the aCGH technique to target different genes/chromosomal regions for diagnosing different DD disorders. We believe that our counting LDTs or FDA-regulated testing kits is appropriate. We agree that, if a whole genome aCGH/microarray test is used, it may detect multiple microdeletion syndromes at the same time. However, not every laboratory uses the whole genome approach. They may do focused/targeted aCGH/microarray testing. As what we identified from the GTR dataset (also confirmed by the clinical investigators from Penn Medicine who work on this Tech Brief), both targeted and whole genome CMA tests may be used for microdeletion syndromes. Not every targeted aCGH/microarray test is intended to diagnose Williams syndrome or velocardiofacial syndrome.

Commentator & Affiliation	Section	Comment	Response
	Findings	Table 3 contains important problems: Rett syndrome is chiefly caused by mutations in MECP2. Why is this gene not listed for Rett syndrome (but for the differential diagnosis in Angelman syndrome??). For Smith-Magenis syndrome and velocardiofacial syndrome and Williams syndrome a CMA is the appropriate first test, it is the diagnostic test. Only for Smith-Magenis may RAI1 sequencing be the next step. FBN1 has nothing to do with velocardiofacial syndrome. And while elastin in deleted in Williams syndrome, ELN1 sequencing is not indicated in Williams syndrome.	MECP2 was mistakenly typed as MEF2C in the previous draft. We have corrected the typo and thank the reviewer for catching it. We double-checked the GTR data and confirmed that MECP2 was indeed listed as the genetic target for 6 Angelman syndrome-related tests. Some literature (e.g., Watson et al. J Med Genet 2001;38:224-228 doi:10.1136/jmg.38.4.224) has linked Angelman syndrome to MECP2. So we deem it appropriate to include those 6 tests in this report. FBN1 tests are typically used for diagnosing Marfan syndrome. We suspect that some laboratories registered these tests with GRT under the velocardiofacial syndrome category for the purpose of differential diagnosis. Based on the reviewer's comment, we decided to remove these FBN1 tests from the velocardiofacial syndrome category for this Technical Brief. We also re-evaluated tests under other DD categories and removed those most likely registered for the differential diagnosis purpose. ELN has an established association with Williams syndrome (http://ghr.nlm.nih.gov/gene/ELN). We deem it appropriate to include ELN tests for Williams syndrome in this report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Findings	The findings are clearly presented and comprehensively outlined in clearly constructed tables. As mentioned before, linking these findings to the guiding and framework questions would provide a clear structure for the reader.	We appreciate the reviewer's comment on the general quality of the report. We have revised the report to link the findings to guiding questions. In addition, we removed the evaluation framework questions to avoid potential confusion with the guiding questions.
Peer Reviewer 3	Findings	The authors highlight the absence of RCTs and general lack of important data that would move the field forward. Specific points: 1. P14 line 3 Comparators for Angelman not appropriate – they are tests of ID. Specific point is not using methylation for UPD – please note that note aCGH will miss UPD.	We have removed the comparators that caused the concern.
	Findings	2. Table 1 is of limited utility – number of tests for ASD – what does that mean, ditto for Angelman, can't assess analytic quality	We have strived to provide useful information in this Technical Brief. As we explained in the report, the number of tests for ASD or Angelman syndrome listed in Table 1 means the "Number of Tests Identified" from GTR (i.e. identified LDTs clinically available in the U.S. for assessing the disorder). Additionally, Table 1 is not intended to report findings on "analytic quality." We address the analytic validity issues in a later section in the report.
	Findings	3. Table 2 – Southern blot not mentioned for Fragile X. This is very important for both females and for males with mosaicism.	We added Southern blot to Table 2 as the reviewer suggested.
	Findings	4. Specific syndromes such as Angelman syndrome are listed, but not others (15q11-q13 duplication, 16p11.2 CNV, etc) are not listed which are more common.	The scope of the Technical Brief was determined based on key stakeholders' input. We include some less common syndromes largely because they are of more interest to the stakeholders.

Commentator & Affiliation	Section	Comment	Response
	Findings	5. Table 3 – Angelman syndrome includes non-AS genes in it (e.g. CDKL5) Rett syndrome has genes that are not Rett syndrome genes and doesn't have the Rett syndrome gene, MECP2 - this is because these are listed as something that is run when AS or Rett is suspected - the review should state that it would be useful to refer to these as tests to be done after AS or Rett is excluded if that is what is suspected clinically and discuss whether to do a full panel or staging (e.g. MECP2 first and other genes for other disorders if no MECP2 mutation).	All data we reported in Table 3 were collected from GTR. These data were voluntarily submitted by laboratories. As we discussed in a response to a similar comment by another reviewer, some of the single-gene tests (e.g., CDKL5 testing for Angelman syndrome and Rett syndrome) might be submitted by labs for the purpose of differential diagnosis. We have reevaluated all the single-gene tests included in the previous draft report and removed those used for the differential diagnosis purpose. MECP2 was not included in the previous draft report due to a typo. We have corrected this typo and the gene is now listed in Table 3.
	Findings	6. The review is confusing in terms of presentation of phenotype vs genotype first approaches and needs to emphasize that one may not replace the other. Both are different levels of diagnosis that have different implications for treatment.	We revised the content regarding phenotype vs. genotype-oriented diagnosis and moved up the discussion to the Background section.
	Findings	7. Table 6-only data on number of subjects and what was examined, but there is no mention of the results (at point of referencing this would be useful to reference results such as in Table 8).	This report is a Technical Brief, not a systemic review. The purpose of the report is to identify potential evidence gaps. It is not intended to review or evaluate the results of the studies identified. Reporting the findings of these studies is beyond the scope of work.
	Findings	8. Listing as a PCR test isn't useful without a detection method – e.g. PCR with sequencing or PCR for number of FMR1 repeats	We revised Tables 4, 5, and 6 in the previous draft (i.e., Tables 4, 6, 8 in the revised report) to provide additional information on the PCR method used in each study. To make Table 2 easier to read, we report PCR tests in one column. Detailed information on the PCR methods used in each test is provided in Appendix D.

Commentator & Affiliation	Section	Comment	Response
	Findings	9. P33, line 13 may be incorrect - We did not identify any empirical study focusing on ethical or legal issues regarding genetic testing in the context of DD care. (see Tabor H et al, for example).	The reviewer did not provide a specific reference for the study by Tabor H et al. Our search identified 3 studies first-authored by Tabor HK within this Technical Brief's search period (PMID: 22532433; PMID: 22038764; PMID: 17873651). These 3 articles address ethical implications of genetic analyses (e.g., whole exome sequencing, aCGH) for Miller syndrome or genetic research. They do not address a DD condition within the scope of this Technical Brief. Based on the reviewer's comment, we revised the report to make following clarification: "We did not include studies that provide general discussions of ethical issues that may apply to DDs and non-DDs."
	Findings	10. Listing recent guidelines only omits important guideline still in place by American Academy of Pediatrics	Given how fast genetic research and testing methods for DDs changes and incorporating some key informants' input, we deem it appropriate to focus on guidelines published in the most recent 5 years. However, based on the reviewer's comment, we added a caveat in the report to remind readers that older guidelines may also provide useful information.
Peer Reviewer 4	Findings	1) Description of proposed interventions: a) Appendix B provides a concise and clear description of the methodologies/technologies used. Table 2 provided information on availability; the participation in PT/lab exchange was an important point.	We thank the reviewer for the comment.

Commentator & Affiliation	Section	Comment	Response
	Findings	b. On page 12, lines 47-52, you note that the offering of multiple tests using different methods might represent newer versions of tests. I would think this unlikely as older versions are generally retired. However, different purposes (e.g., diagnosis, screening) may translate to different tests that differ in cost (lower for screening), clinical sensitivity (higher for diagnosis), turn-around-time, and other factors. Differences in gene markers could also relate to test purpose, but indications for use is key information (informative if this is generally missing in GFR).	We thank the reviewer for the comment and have removed the content that caused confusion.
	Findings	c. Table 3 is clear and informative, but raised questions for me. First, it seems that there could be two 'categories' of tests here: - the syndromes that represent more classic diagnosis based on known causative genes (e.g., FMR1, maternal del in UBE3A), chromosome abnormalities, or genes with known association with a phenotype (CDKL5). - the broader DDs of ID, ASD and GDD for which clinical and developmental assessment are an important part of clinical diagnosis and the goal of genetic testing is to identify a genetic etiology or perhaps add information that is/is not consistent with the clinical diagnosis.	We have revised the report to differentiate ID/ASD/GDD from the 8 DD syndromes included in the report. We think it is still reasonable to include the test information for these DD conditions in a single table. This helps avoid creating too many tables in the report.
	Findings	d. The last column in Table 3, tests for which specific genes are not reported is striking – not sure what clinicians would do with such results? Good example of potential risk of testing?	We revised the heading of the column that caused the confusion. Now it reads: "Numbers of Tests Analyzing a Chromosomal Region or the Whole Genome or Exome."

Commentator & Affiliation	Section	Comment	Response
	Findings	2) Evidence map a. On page 18, lines 31-33, you state that results can be interpreted as addressing either analytic or clinical validity. In the way I think you mean it, this is pretty unlikely, especially when 9 of the 21 studies in Table 5 were focused on FraX syndrome and two more on Smith-Magenis. This table could be made clearer by designating what was reported using standard terms. - For example, Truong et al. 2008 describe a validation of a Q-PCR method that they report can identify deletions (SMS) and dup17p11.2 (duplication syndrome) as confirmed by comparison to two other methods (FISH, MLPA). So analytic validity was addressed, though the specifics did not jump out at me. The non-overlapping normal, deleted and duplication ranges for RAI1 do provide a reportable range, but also provide evidence of clinical validity, as SMS duplications and dup17p11.2 can be distinguished from each other and from normal – estimates of clinical sensitivity and specificity should be possible. - Lafauci 2013 is another good example. As you noted analytic validity (at least repeatability) was established using control samples (though again I saw no specifics). However, the normal distribution of FMRP (gene expression) in males/females is provided, so along with distributions for fraX males, mosaic/non-mosaic males, premutation females, and fraX females this would allow computation of estimates of clinical sensitivity/specificity and ppv/npv for identification of each.	Based on several reviewers' comments, we revised the paragraph and separated the analytic and clinical validity studies. Now, Table 5 focuses only on analytic validity studies. The small number of clinical validity studies are discussed in body text. The Truong study validated a new PCR test for measuring the copy numbers of the RAI1 gene. The repeatability, precision, and reported range of the test were reported. According to the authors, the study validated the test "by conducting a blinded study with samples that have a known deletion or duplication of RAI1" and verified the results "using FISH and multiplex ligation dependent probe amplification." The Lafauci study assessed an immunoassay based on a Luminex (Austin, TX) platform that detects fragile X mental retardation protein level. This study purportedly explores the possibility of developing a fragile X test for newborn screening. Both the Truong and Lafauci studies focus on assessing the test's ability to detect an analyte instead of connecting the analyte to the clinical disorder. We deem both analytic validity studies.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 1	Summary and Implications	The main concern is that insurance companies may interpret this report as justification for denying genetic testing to patients with ID and ASD, as the benefit of doing the testing has "not been proven" in a comprehensive review of the literature.	We understand the reviewer's concern. This Technical Brief is intended to identify published evidence that may address the clinical utility of genetic tests, and we strived to make our literature search/screening as thorough as possible. The gaps in published evidence that we have identified do not necessarily suggest genetic testing does not have benefits or value. Instead, it suggests more published evidence is needed to demonstrate genetic testing's value/benefits for assessing DDs.
	Summary and Implications	One issue which has not been truly addressed is WHO is actually ordering and interpreting the testing and its results. If expensive testing is ordered, but it is not the correct test to order and comes back "negative", than of course the usefulness of the testing would be minimal. It would be interesting to evaluate the outcome in terms of medical benefit for patients who had actually been seen by a geneticist and had appropriate testing and genetic counseling performed.	One of the report's guiding questions is: Who are the providers ordering the tests and using their results? Our searches did not identify any study that reported real-world data to address this question. We agree with the reviewer that the effectiveness of genetic testing may be affected by who orders the tests or who interprets the test results. Our literature search/screening did not exclude studies that compared the effectiveness (accuracy, impact on health outcomes, etc.) of genetic testing ordered or interpreted by different clinicians (e.g., primary care physicians vs. geneticists). If this type of evidence existed, we would have captured it in Tables 4–6.
KI Reviewer 2	Summary and Implications	The conclusion appears to be that the evidence base is inadequate and more focused research needs to be done. This is again not a surprise since those who have tried to develop evidence based guidelines are already aware of this. The brief therefore is a well documented summary of what is already widely suspected.	We appreciate the reviewer's comment.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 3	Summary and Implications	On page 33, the paragraph beginning on line 41 discusses the need for well-designed RCT or non-RCT to answer the clinical utility question. While one could hardly disagree with this statement, a relevant question is how often these types of studies have been applied to other diagnostic modalities outside of the realm of genetics (imaging for example). As an example Brain MR scanning, an expensive modality, is frequently applied to a variety of conditions including the conditions chosen in this report. As noted the diagnostic yield of this test in the context of ID, ASD, GDD is very low, yet it is not scrutinized in the manner that genetic diagnostics are. I'm not arguing that genetic tests should get a free pass, but what I am saying is that genetic diagnostics are begin treated in an exceptional manner compared to other diagnostics.	We agree with the reviewer that there are many challenges for doing RCTs or even non-RCTs to directly evaluate genetic testing's impact on health outcomes. We had discussed these challenges in the previous draft. Based on the reviewer's comment, we further expanded the discussion. However, regardless of the challenges, it is still feasible to design and execute clinical utility trials for certain tests and disorders; therefore, we encourage researchers to make efforts in that direction.

Commentator & Affiliation	Section	Comment	Response
	Summary and Implications	Page 34 paragraph beginning on line 21 reiterates the issues related to diagnostic yield and ends with the statement, "Improved diagnostic yield may not necessarily lead to a positive change in clinical management or in health outcomes." This is stated with a negative bias which for the reasons articulated above regarding condition specific management should be more balanced. As an example, in table 7 there is reference to a new CNV deletion 16p11.2 as a significant cause of ASD. Study of individuals with this deletion has identified that in addition to ASD these children are at increased risk for developing obesity. Recommendations for weight management and nutrition counseling are now included in the management sections of the GeneReview article for this condition (http://www.ncbi.nlm.nih.gov/books/NBK11167/). Claims of utility cannot be made as there are no studies demonstrating that these effectively prevent obesity and is co-morbidities. However, one of the conditions explicitly considered, Prader-Willi syndrome is also predisposed to abnormal eating behaviors and severe obesity. Intensive weight and behavioral management has been shown to be very effective in preventing obesity (including in a handful of small prospective case-control studies, e.g. J Pediatr Endocrinol Metab. 2008 Jul;21(7):651-5. Successful early dietary intervention avoids obesity in patients with Prader-Willi syndrome: a ten-year follow-up. Schmidt et al.). This is not mentioned even though most would argue that prevention of severe obesity confers positive clinical utility. There are many other examples of newly discovered copy number variants identified through aCGH and CMA that are associated with other medical conditions that have resulted in condition specific medical management recommendations. I am not arguing that there is definitive evidence of utility, rather that the negative tone of the discussion of the lack of evidence needs to be tempered.	We agree with the reviewer that improved diagnostic yield may have impact on patient management. We have revised the sentence that caused the concern. Now the sentence reads: "Improved diagnostic yield may have an impact on patient management because some condition-specific management (e.g., obesity prevention programs for DD patients at higher risk for developing the comorbidity) cannot be initiated without a specific diagnosis (e.g., an etiologic diagnosis revealing a patient carries a genetic variant associated with obesity)."

Commentator & Affiliation	Section	Comment	Response
	Summary and Implications	Page 35 in the Limitations section where there is discussion of inclusion of a limited number of genetic conditions, it should be noted (as I have emphasized above) that no genetic conditions with effective treatments were included which limits the assessment of clinical utility.	As we note in the report, the scope of this report, including the 11 DD conditions, was determined based on the interest of key stakeholders. The report is not intended to include every DD condition. The findings of this Technical Brief only apply to the DD conditions included in the report. We have revised the Summary and Implications section to clarify this point.
	Summary and Implications	Page 34. American College of Medical Genetics and Genomics	We have corrected the name of ACMG. We thank the reviewer for catching the error.
	Summary and Implications	Page 36. For new technologies no gold standard exists	We revised Table 9 based on this comment. Ideal evidence for addressing clinical validity is now described as "cohort studies that evaluate the test's diagnostic accuracy using phenotype or other acceptable diagnostic standards as reference methods."
KI Reviewer 4	Summary and Implications	It would seem that given all the work put into this project, a stronger statement should be made about the lack of real data to support the growth of testing.	In this section, we summarized the evidence gaps for addressing the clinical utility of genetic tests for assessing DDs. We also discussed what kind of future studies are needed to fill the gaps. We think the message about the current evidence status is quite clear. We decided not to make a potentially controversial statement such as "lack of real data to support the growth of testing," because researchers, test developers, and policymakers may not agree on what constitutes "real data" for supporting genetic tests' clinical utility.
Peer Reviewer 1	Summary and Implications	This section appropriately summarizes what was found.	We appreciate the reviewer's assessment.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Summary and Implications	The summary is an accurate summary reflection of the findings. Presenting the summary using the framework of the evidence chain questions would be helpful for the reader. Indeed, this is mentioned as being presented in Table 10, but there is no Table 10 in the document.	We appreciate the reviewer's comment. We have corrected the typo with the table number. We also removed the evaluation framework and the evidence chain questions to avoid potential confusion caused by that content.
Peer Reviewer 3	Summary and Implications	Summary and implications was a strength of the paper and would be strengthened when discussing the limitations to point out the errors (potential of interpretation) if taken at face value. The review could further talk about the rapid evolution in the field especially in terms of WES and CGS for both SNVs and CNVs and small indels of clinical significance.	The reviewer's comment on "errors (potential of interpretation) if taken at face value" is not specific enough for us to take an action. Regarding the suggested discussion on WES and CGS, see our response to next comment.
Peer Reviewer 4	Summary and Implications	1) On page 34, lines 16-20 you note that most of the 21 AV/CV studies 'were intended to validate the performance of a newly developed test'. Well, that is pretty much the only way such data can get into publication and that is what you have in most cases (unless there are data from PT). The next step (which is outside the scope of this Tech Brief I believe) would be to check if any of these validation studies relate to tests currently in use. Relevant studies would need to be evaluated for study design/quality issues and strength of evidence before concluding that further research/validation is needed.	We revised the sentence that concerned the reviewer to avoid potential negative interpretation of our view about the value of the AV/CV studies. However, we still contend that the findings of these studies on new tests need further validation by future research. We agree with the reviewer that systematic review of these validation studies' design/quality/strength of evidence is beyond the scope of this Technical Brief.
	Summary and Implications	2) In my view, the number of CV studies of 0 in Table 9 is questionable and the starred footnote is a bit confusing. I do not mean to suggest that information on AV and CV is widely available, especially in these new technologies. However, an important role of these Tech Briefs in my opinion is an accurate assessment of what might be out there.	We revised the sentence that concerned the reviewer to avoid potential negative interpretation of our view about the value of the AV/CV studies. Because of the revision we made, the number of CV studies in Table 9 has changed. We also added a note under the table to explain why we do not expect a big number of studies in that category.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 1	Next Steps	I agree that better research is needed to prove the benefit of genetic testing for patients with ID and ASD. I think that it might help to look at different tests and diagnoses more individually, and also compare the utility of who is ordering and interpreting the tests.	We agree with the reviewer and have revised the section to emphasize the value of future studies that compare the utility of who is ordering and interpreting the tests.
KI Reviewer 2	Next Steps	The feasibility of the types of studies suggested needed to be examined further. The test environment is changing so quickly that the longitudinal studies desired might never happen before the "next tests" are out. It would have been really helpful if there was a suggestion of any type of relevant research that could be done in a short time frame.	Per the reviewer's suggestion, we expanded the discussion on practical challenges of conducting longitudinal studies. At various sections of the report, we discuss how to use different types of evidence to build a chain of evidence to address the clinical utility when the ideal type of evidence is not available. Therefore, we elect not to further elaborate the evidence issue.
KI Reviewer 3	Next Steps	Other than the 'standard' more and better evidence is needed and that funding is needed for such studies (thank you for including that!!) there are no recommendations that would meet the language in the directions to be "…as specific as possible…".	In the report (see Table 9), we specifically describe current evidence gaps for addressing the clinical utility, clinical validity, and analytic validity of genetic testing. We also offered specific opinions about the types of evidence that are ideal or helpful for addressing those issues (also see Table 9).
KI Reviewer 4	Next Steps	Getting the word out that the Emperor is wearing no clothing	Once this report is finalized, it will be posted on the AHRQ's Web site. The agency may also use other channels to disseminate information.
Peer Reviewer 1	Next Steps	The issues for future research are well delineated.	We appreciate the reviewer's assessment.
Peer Reviewer 2	Next Steps	Of note: Table 9 is particularly helpful in identifying the gaps in evidence and outlining the next steps as specific next steps are not listed. These should be elucidated beyond the sentence in the summary, perhaps building off Table 9 and incorporating perspectives on the newer genetic testing approaches not addressed in this brief.	We appreciate the reviewer's comment on the utility of Table 9. This table elucidates current gaps in evidence for addressing the clinical utility of genetic tests. The identified evidence gaps overarch both newer and older testing approaches. We feel there is no need to call out newer testing approaches. In our opinion, the evidence gaps for both newer and older testing approaches should be addressed in the future.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Next Steps	Important next steps are described including databases that are more useful than GTR. They may want to more specifically recommend revision of GTR (or its removal). The review could more further talk about the rapid evolution in the field especially in terms of WES and CGS for both SNVs and CNVs and small indels of clinical significance.	In the report, we have discussed potential limitations of the GTR database. We also communicated with the GTR staff for any concerns or specific suggestions on technical matters. We elect not to discuss whole exome sequencing (WES) and comparative genome sequencing (CGH), in this section because the data that we collected and analyzed do not support an insightful discussion on the subject. The "next-steps" discussions currently included in the section are all based on the findings of the report.
Peer Reviewer 4	Next Steps	The next steps seemed to be integrated into the above section (e.g., page 34, lines 55-56) and a bit hard to identify. I think a separate section with bullet points would be very useful, as next steps would seem to be important take-home messages to sum up this review.	The suggestions regarding "next steps" in research are provided immediately after the discussions of each category of evidence gap (i.e., RCTs and non-RCTs, clinical or analytical validity studies, diagnostic yield studies). Moving those suggestions would make the discussions out of context. So, we decided not to move those suggestions together into a new section.
KI Reviewer 1	Clarity and Usability	The report is well written. The tables are very detailed and somewhat confusing. There is insufficient explanation as to how the tests should be chosen and in what order.	We appreciate the reviewer's comment on the general quality of the report. The comment on the tables are not specific enough for us to take an action. This report includes a broad range of development disorders, and clinicians/geneticists do not always agree on the diagnostic approaches for these disorders. Instead of offering too much discussion of the diagnostic approaches that reflects our own viewpoints, we summarized consensus statements or recommendations (Table 7) from clinical practice guidelines regarding how to use genetic tests in DD diagnosis. We are hoping these guidelines provide more balanced views.

Commentator & Affiliation	Section	Comment	Response
KI Reviewer 2	Clarity and Usability	Structure is fairly good but could be slightly more clear. Would have liked to see the phenotype/genotype question addressed early and not leave such important points for so late in the document. Also would like to have seen bulleted implications and suggestions.	We are hoping the revisions we have made based on the reviewer's previous comments make the report clearer and easier to read. The phenotype/genotype discussion has been moved up as the reviewer suggested. However, we decided to discuss those implications and suggestions separately in different paragraphs where there is a context for the discussions.
KI Reviewer 3	Clarity and Usability	The report is well organized, clearly presented and easy to read. As to the usability, this is tied to some of the weaknesses pointed out in the prior sections. These would need to be addressed for the report to be optimally used.	We thank the reviewer for taking time to review this report and providing insightful comments.
KI Reviewer 4	Clarity and Usability	Generally written clearly and with a structure that allows for understanding and flow. This is a good step in "setting the baseline" and hopefully will push for more studies of actual outcomes.	We appreciate the reviewer's support.
Peer Reviewer 1	Clarity and Usability	Overall, the report is well structured. The main issue, in particular difficulty in showing a positive effect of genetic testing on long term medical outcome, is clear.	We appreciate the reviewer's assessment.
Peer Reviewer 2	Clarity and Usability	Overall the clarity is good, but additional structure for the reader, specifically in relation to the guiding questions and the questions forming the framework for establishing the chain of evidence, would enhance the brief considerably. The description of the methodology is very clear. Overall, the brief is very useable with excellent and comprehensive tables for understanding the range and breadth of genetic testing for DDs.	We appreciate the reviewer's comment on the general quality of the report. We have revised report to link methods and findings to the guiding questions. We also removed the key questions associated with the evaluation framework to avoid potential confusion with the guiding questions.
Peer Reviewer 3	Clarity and Usability	The report is useful at the higher level but many of the tables serve to demonstrate how much needs to change to improve the organization of the field from the perspective of what tests are offered and why.	We thank the reviewer for taking time to review this report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Clarity and Usability	Overall, this report is organized and I noted main points throughout. I had a little trouble finding the answers to all Guiding Questions, though I may have missed them. I think perhaps some additional work on pulling the conclusions together using bullets or the Guiding Questions would be very helpful and would lead in to a final section on Next Steps.	Thank you for your feedback. We have addressed your specific comments concerning the Guiding Questions and Next Steps in those sections. We have revised the Finding section to match the content to relevant guiding questions. Given that this Technical Brief addresses so many guiding questions and sub-guiding questions, we feel it is not effective to touch every question again in the Summary and Implications section. We choose to focus on discussing main themes about the identified tests and evidence gaps.