Assessing the analytic validity of molecular testing for Huntington disease using data from an external proficiency testing survey

Glenn E. Palomaki, PhD[1] and C. Sue Richards, PhD[2]; on behalf of the CAP/ACMG Biochemical and Molecular Genetics Resource Committee

Purpose: Documenting high analytic validity of the molecular diagnostic test for Huntington disease is important because of counseling implications. This dominantly inherited adult onset disorder (prevalence of three or more per 100,000) is characterized by chorea, ataxia, and personality changes. The molecular basis is excessive CAG repeats in the *HTT* gene.

Methods: External proficiency testing survey results for Huntington disease were extracted (2003–2010). Analytic interpretations and CAG repeat lengths were compared with published performance criteria.

Results: Between 2008 and 2010, 33 US participating laboratories reported clinical test interpretations. Analytic validity was high (sensitivity: 99.5%, 95% confidence interval: 97.1–99.9%; specificity: 99.2%, 95% confidence interval: 97.1–99.9%). Repeat length errors

occurred in 2.6% (95% confidence interval: 1.8–3.8%) of 1,060 allelic challenges, with most being minor or from a single participant. Past performance (2003–2007) was similar. The 23 international participants had more total repeat length errors (17.5%, 95% confidence interval: 14.6–20.7%). Further analyses indicated that assessment criteria can be relaxed without jeopardizing analytic validity.

Conclusion: Analytic validity is high for Huntington disease testing among US laboratories. International survey participants had lower analytic validity and a higher proportion of poorly performing laboratories. The reasons for this are unclear.

Genet Med 2012:14(1):69–75

Key Words: analytic validity; external proficiency testing; Huntington disease; quality assurance

INTRODUCTION

Huntington disease (HD) is a devastating neurologic disorder that affects approximately 3–7 per 10,000 in individuals of western European ancestry.[1,2] Disease characteristics include chorea, ataxia, and personality disorders. A majority of patients have adult onset, with symptoms beginning around age 40 years. Juvenile onset occurs in approximately 5–10% of cases. HD is a dominant genetic disorder, resulting from expansion of a CAG triplet repeat in exon 1 of the *HTT* gene located on 4p16.3. If the number of CAG repeats is 40 or higher, the disorder is considered to be fully penetrant. Among patients with and without clinically defined HD, specific size categories of repeats correlate with disease phenotype. In the clinically normal population, 25 or fewer repeats are associated with a stable repeat size, but between 26 and 35, the triplet repeat may increase in size in the next generation. Individuals with between 36 and 39 repeats may or may not develop symptoms ("reduced penetrance").[3] Generally, the larger the repeat size, the earlier age of onset of symptoms. However, prediction of age of onset based on repeat size is inaccurate and, thus, not used clinically. Because the CAG expansion for HD is relatively small, accurate sizing of the repeat is essential for accurate classification and prediction of disease state. Therefore, laboratory testing for HD needs to be

highly accurate and associated with an appropriate analytic interpretation based on repeat length. Questions have been raised as to whether laboratories in the United States are performing at sufficiently high standards.[4,5]

Excessive CAG repeats (located in the coding region of the *HHT* gene) result in an abnormal protein product, Huntingtin, with an aberrant function. The repeating units of CAG encode glutamine, resulting in very long stretches of glutamine in the protein product. These polyglutamine tracts result in abnormal protein-protein interactions and produce accumulations of aggregate proteins in neural tissues. These inclusions are hallmark features of the disorder. Thus, HD is in a class of trinucleotide repeat disorders that are often referred to as polyglutamine disorders, including spinalbulbar muscular atrophy, multiple spinocerebellar ataxias, and dentatorubral-pallidoluysian atrophy. Some of these disorders have overlapping symptoms that can be confused with HD, particularly dentatorubral-pallidoluysian atrophy. Therefore, diagnostic testing of specific CAG repeats is useful in the differential diagnosis of patients.

CAG repeats can be easily detected by polymerase chain reaction (PCR) assays coupled with some form of electrophoresis to size the products. In HD testing, primers are designed to match the flanking region of the CAG repeats to generate

ORIGINAL RESEARCH ARTICLE

an allele product, and the size of the allele is determined by electrophoretic mobility. As a result, many clinical molecular genetic laboratories are able to provide genetic testing for HD. Allele sizes >100 repeats may be seen in the juvenile form; this requires a Southern analysis, which is restricted to fewer laboratories. In addition, Southern analysis may need to be performed in cases of homozygosity of a normal size allele. As an alternative approach for differentiating two normal alleles, some laboratories use primer sets, which include the adjacent CCG repeat, located 3′ to the CAG repeat. Generally, HD testing is performed for confirmation of diagnosis of a symptomatic patient or for presymptomatic testing in family members of an affected individual. However, prenatal testing and preimplantation genetic diagnosis testing are also available for these families. Pretest counseling, neurologic examination, and informed consent are obligatory, and laboratories generally act as gatekeepers. The incidence of depression in presymptomatic and symptomatic patients with HD is more than twice that of the general population and a higher rate of suicide has been reported.[6]

Recognizing the significance of this devastating genetic disorder and the importance of accurate testing results for sizing of these HD alleles and interpreting them, the College of American Pathologists (CAP) and the American College of Medical Genetics (ACMG) began offering external proficiency testing for this disorder in 1996. The oversight of this testing program rests in the CAP/ACMG Biochemical and Molecular Genetics (BMG) Resource Committee. The BMG Resource Committee is composed of CAP- and ACMG-appointed experts in the areas of clinical genetics testing. The Committee members select sample challenges from validated reference materials with known genotypes (Coriell Institute for Medical Research, Camden, NJ) that are sent to participants enrolled in the CAP/ ACMG proficiency testing survey. The laboratories perform the analysis and report their results and interpretation back to CAP, which performs defined analyses and forwards the results to the Committee members for review and discussion. Three challenges are provided twice a year, and all CAP accredited laboratories that provide HD testing must be enrolled in these surveys. Samples are extracted DNA and are shipped at room temperature for overnight delivery. This results in a stable sample material for testing. Thus, sample degradation resulting in incorrect analytic results should not be an issue. In the past, this survey was graded solely on the interpretation when there was 80% consensus among participants. In the current analysis, we also include simulated grading based on allele sizing, using criteria from the ACMG Huntington Disease Standards and Guidelines,[3] which is an update to an earlier document on the same topic.[7]

Grading of the allele size began in 2011 and requires that laboratories both accurately size and interpret these alleles. Herein, we analyze CAP/ACMG HD survey challenges from 2003 to 2010 to answer four main questions: (1) among current participants in the Survey, how many clinical tests are performed and what is the turn-around time? (2) What is the recent analytic performance of participants in the United States? (3) How does that performance compare to international participants? and (4) Has the performance of the US participants changed over time? We also explore whether errors tended to be randomly distributed or localized to a few participants, were associated with test methodology, or were associated with the numbers of samples tested monthly. Finally, we examined the precision profile of repeat length in US participants to determine whether existing guidelines for sizing precision were appropriate. This report is one in a series of articles that will actually use CAP/ACMG proficiency testing data to address the quality of genetic testing in US laboratories.

MATERIALS AND METHODS

The ACMG and the CAP jointly sponsor an external proficiency testing module for HD under the auspices of the CAP/ACMG BMG Resource Committee (the Molecular Genetics Laboratories-BMG Survey). Participants receive three samples twice a year for a total of six challenges per year. Data were only available from 2003 to 2010, inclusive. In addition to quantifying the number of CAG triplet repeats for each allele and providing a clinical interpretation based on the largest sized allele, each participant is asked to provide additional data including (1) whether they provide clinical services (as opposed to research testing), (2) the numbers of tests for HD performed each month, and (3) the average turn-around time for clinical testing. CAP staff classified participants as being located within the United States or international, based on the shipping address used to send sample challenges. The authors did not have access to the actual locations or the identification of participants. For each participant, an average number of samples per month and average turn-around time were computed using all provided responses.

Overall error rates, types of errors, and analytic sensitivity and specificity were documented. Analytic sensitivity was defined as the proportion of individual participant challenges of 43 repeats or more that were reported as being at least 40 repeats (consistent with complete penetrance). Analytic specificity was defined as the proportion of individual participant challenges of 23 repeats or fewer that were reported as being no more than 26 repeats (unaffected, normal allelic size). The larger median number of repeats for each challenge (consensus result) was used to determine the appropriate interpretation, based on the ACMG Laboratory Standards and Guidelines for HD.[3] An error in analytic interpretation occurs when the participant's reported interpretation differs from the ACMG recommendations of "normal allele, no risk" for ≤26 repeats, "normal mutant allele, unaffected" for 27 to 35 repeats, "expanded allele, variable penetrance" for 36 to 39 repeats, and "expanded allele, complete penetrance" for ≥40 repeats. The ACMG recommended precision guidelines for HD allele sizing were used to grade reported repeat lengths. For ≤43, 44–50, 51–75, and >75 repeats, the allowable range was ±1, ±2, ±3, and ±4 repeats, respectively.[3]

A precision profile was created for US participants by using a two-pass three-standard deviation trimming algorithm and then plotting the coefficient of variation (trimmed standard

deviation/trimmed mean × 100) versus repeat length. A second-order polynomial was fitted to provide an estimate of the precision profile. Converting a coefficient of variation to a repeat length was the inverse of this process. Comparison of rates was by an exact two-tailed test. Confidence intervals (CIs) were computed using the binomial distribution. Tests of trends for continuous data were performed using linear regression and test of the slope equal to zero. Significance was at the $P = 0.05$ level. Analyses were performed using True Epistat (Richardson, TX). Graphics and selected statistics performed using GraphPad Prism V5 (La Jolla, CA).

RESULTS

HD testing turn-around time and numbers of samples tested
Between 2008 and 2010 inclusive, the six sets of three sample challenges were distributed to 35 US participants and 24 international participants. Three additional participants (one international and two US participants) were excluded from consideration, as they took part in only one distribution occurring in 2008. The 33 US participants collectively processed an estimated 423 samples per month, averaging 12.8 samples per laboratory (range: 1–92). The median turn-around time for clinical service was 14 days (range: 2–25). International participants collectively processed an estimated 107 samples per month, averaging 5.1 samples per laboratory (range: 1–20). The median turn-around time was 21 days (range: 3–90). All participants reported clinical results. To investigate whether those doing more testing might return results sooner, participant testing numbers were plotted versus turn-around time (**Figure 1**). There was no significant relationship (test of slope, $P = 0.4$). However, turn-around times for US participants were considerably shorter. Of the 32 US participants reporting turn-around times, 31 (97%) reported results in 21 days or less. Among the 23 international participants, 12 (52%) reported results within 21 days (exact two tailed, $P < 0.001$).

Recent analytic performance of laboratories located in the United States
Table 1 lists the sample challenges from 2008 to 2010 along with the corresponding analytic interpretation for HD. In all instances, the median consensus of the participants confirmed the expected genotype of the challenge. The last two columns show the data needed to compute analytic sensitivity and specificity. Overall, there were seven challenges associated with a high number of repeats (>42) that were considered suitable for computing analytic sensitivity. Only one analytic interpretation of "normal allele, no risk" for challenge 2010–15 was incorrect. The participant reported identifying only 16 repeats, whereas the consensus result was 59 repeats with an interpretation of "expanded allele, complete penetrance." This error was likely due to allele dropout, as this participant only identified the normal size allele. Overall, the estimated analytic sensitivity for US participants was 99.5% (95% CI: 97.3–99.9%).

Nine challenges were associated with a low number of repeats (<24) consistent with a normal phenotype, suitable for analysis
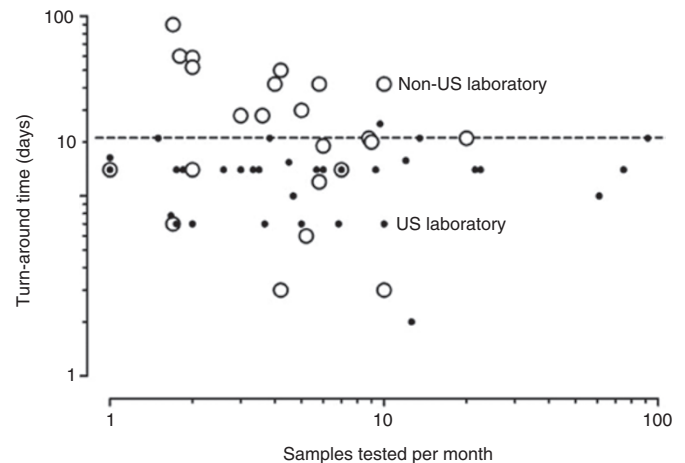
Figure 1 Number of Huntington disease tests performed per month versus turnaround time. Closed circles indicate the information from 32 laboratories located in the United States (one did not report turn-around time), whereas the 23 open circles indicate information from participants located outside of the United States.

of analytic specificity. One interpretation of "normal mutable allele, unaffected" for challenge 2010–13 (**Table 1**) was incorrect; the participant reported identifying 29 repeats, whereas the consensus was 20 repeats. The correct interpretation was normal allele, unaffected. A second interpretation error was less obvious. For challenge 2008–05, the consensus was 22 repeats, but a participant reported finding 29 repeats. That participant reported a "normal allele, no risk" interpretation, but according to current guidelines, the interpretation for 29 repeats should have been "normal mutable allele, unaffected." Overall, the estimated analytic specificity was 99.2% (95% CI: 97.1–99.9%). The two remaining challenges were both associated with a "normal mutable allele, unaffected" and were not suitable for computing analytic performance. However, six participants correctly measured the number of CAG repeats but reported the interpretation to be "normal allele, unaffected" for both. All remaining participants correctly reported the analytic interpretation to be a "normal mutable allele, unaffected."

In addition to obtaining the correct interpretation, it is important for participants to also obtain accurate estimates of the number of CAG repeats. If each allele was considered to be a separate assessment, a total of 1,060 assessments are available for analysis. For this analysis, all challenges are considered suitable, regardless of the analytic interpretation. Using the definitions described earlier, 1,032 of these assessments were within the acceptable range (97.4%, 95% CI: 96.2–98.2%). A total of 28 repeat length errors were identified (2.6%, 95% CI: 1.8–3.8%). Sixteen of these were within one repeat of the acceptable range. However, 12 were two or more repeats outside of the acceptable range, including four highly discrepant results (five or more repeats outside of the acceptable range). Although some highly discrepant results did not result in an incorrect analytic interpretation (e.g., 82 repeats reported with a consensus of 59 repeats—both reported as "expanded

Table 1 Summary participant results from the Huntington disease (HD) proficiency testing survey between 2008 and 2010, restricted to those located in the United States

Challenge	HD alleles	"Expected" interpretation		Participant interpretation		Analytic performance
		Sensitivity	Specificity	Agreed	Disagreed	
2008–04	21,44	Expanded	22	—	22/22	—
2008–05	15,16	Normal	21[a]	—	—	21/21
2008–06	17,22	Normal	22	—	—	21/22[b]
2008–13	21,44	Expanded	30	—	30/30	—
2008–14	17,17	Normal	29[a]	—	—	29/29
2008–15	18,59	Expanded	30	—	30/30	—
2009–04	16,17	Normal	28[a]	—	—	28/28
2009–05	19,20	Normal	29	—	—	29/29
2009–06	16,64	Expanded	29	—	29/29	—
2009–13	17,29	Normal Mut	24	6 Normal	—	—
2009–14	17,22	Normal	30	—	—	30/30
2009–15	21,44	Expanded	30	—	30/30	—
2010–04	17,29	Normal Mut	24	6 Normal	—	—
2010–05	17,22	Normal	30	—	—	30/30
2010–06	21,44	Expanded	30	—	30/30	—
2010–13	19,20	Normal	27	1 Normal Mut	—	27/28[c]
2010–14	17,22	Normal	28	—	—	28/28
2010–15	18,59	Expanded	27	1 Normal	27/28[d]	—
All (%)					199/200 (99.5)	244/246 (99.2)

CAP, College of American Pathologists; HD, Huntington disease.

[a]Results exempted from CAP grading and were not considered for analysis. [b]One participant reported a repeat length of 29 with an incorrect "normal–no risk" interpretation. [c]One participant reported a repeat length of 29 with a "normal–mutation" interpretation. [d]One participant reported a repeat length of 18 with a "normal–no risk" interpretation.

allele–complete penetrance"), this finding might still indicate an underlying technical problem. Other errors may have been due to allelic dropout and incorrectly reporting a homozygous normal result (e.g., reporting a homozygous genotype of 18,18, but the correct genotype is 18,59). Twenty-seven of the 33 participants had no repeat length errors noted; the 28 errors were from the remaining six participants (18%). **Figure 2a** displays the detailed repeat length grading results for the US participants analyzed.

Stratifying analytic performance for US participants by test methodology and sample volume

All the US participants used a PCR-based methodology to generate alleles followed by either capillary gel electrophoresis (majority of participants) or gel-based electrophoresis for separation of alleles. Only one participant used bisulfite treatment of DNA followed by PCR. Of the 36 allele challenges, this participant had nine sizing errors (25% error rate), representing nearly 1/3 of all sizing errors found for US participants. Were these to be removed, the error rate would be reduced to 1.9% (19/1,024, 95% CI: 1.1–2.9%). This method, which converts cytosine to uracil has been shown to have some limitations, including incomplete conversion and degradation of DNA

during treatment, which can result in problems during PCR.[8,9] In addition, this participant may have been separating alleles on an agarose gel (as described in the original publication) instead of an acrylamide gel, which could also be responsible for sizing errors. Based on these data, bisulfite PCR seems not to be an optimum method for sizing alleles.

It is possible that participants processing a higher number of samples per month might be more proficient and, therefore, have fewer errors in sizing repeat lengths. To examine this, the US participants were divided into two equal groups; the 16 reporting fewer than five samples tested per month versus the 16 participants reporting five or more (one participant did not report numbers of samples). The error rates for these two groups were not significantly different at 2.0% (10/502) and 3.2% (18/558), respectively (two-tailed exact, $P = 0.2$).

Recent analytic performance of international participants

Of the 23 international participants, 16 were from Canada, two from Saudi Arabia, and one each from Hong Kong, Qatar, Singapore, South Korea, and South Africa. During the 2008–2010 time period, the 23 international participants were subject to 612 assessments. **Figure 2b** displays the detailed repeat length grading results for these international participants.
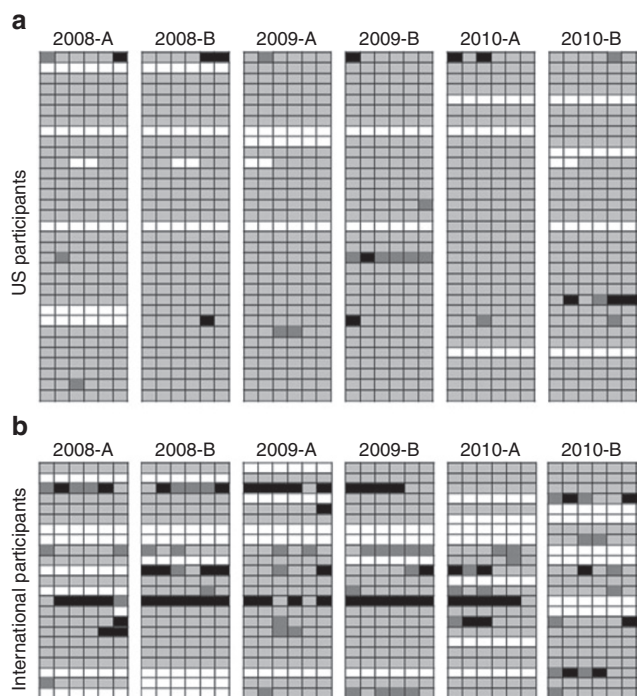
Figure 2 **Individual participant performance for repeat sizing during 2008–2010, stratified by geographical location of the participant.** Each row represents an individual participant set of results, with rows in descending order by average number of samples tested per month. The fill indicates participants' sizing performance for each allele (paired columns indicate a single challenge). Light gray indicates the sizing results are within the acceptable range (see text for definition), dark gray indicates results are outside of the acceptable range by only one repeat, black indicates outside of acceptable range by two or more repeats, and white shows distributions (samples) for which evaluation was not possible, usually because no results had been returned. (**a**) Results from participants located within the United States. (**b**) Results for international participants.

Overall, 107 repeat length errors were identified (17.5%, 95% CI: 14.6–20.7%). This is higher than the 2.6% rate found for the US participants (exact two tailed, $P < 0.001$). Among the 107 errors, 49 were one CAG repeat outside of the acceptable range. However, 45 errors were five or more repeats from the acceptable range (7.3% vs. 5% for US participants, $P < 0.001$). Repeat length errors were identified for 14 of the 23 international participants (61%), significantly higher than the 18% of US participants ($P < 0.001$). Three international participants were especially problematic with, on average, sizing errors on 57% of alleles tested (72 of 126 alleles, 95% CI: 48–66%). In the other 20 participants, the error rate was 7.3% (35/486, 95% CI: 5.1–9.9%). This is still significantly higher than the 2.6% found for US participants ($P < 0.001$). Canadian participants performed similarly to those located elsewhere (18% vs. 6%, exact two tailed, $P = 0.6$).

Comparison with past performance for US participants

Most US laboratories that participated in proficiency testing in 2008–2010 also participated between 2003 and 2007. Performance comparisons for these two time periods are
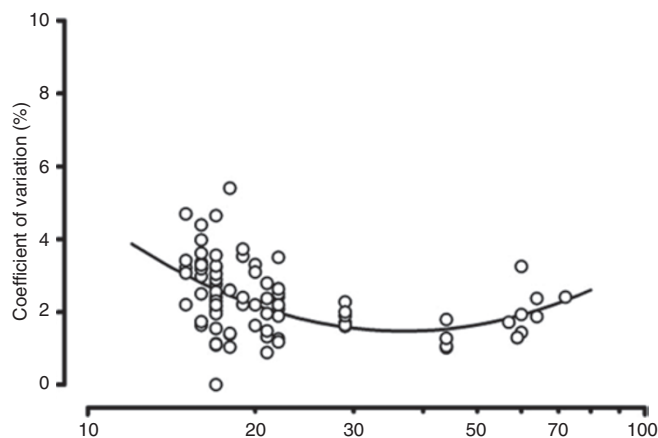
Figure 3 **Precision profile of the trimmed CAG repeat length.** Each allele was treated as an independent result and the reported lengths subject to a trimming algorithm. The trimmed coefficient of variation (CV) (y axis) was plotted versus the consensus repeat length (logarithmic x axis). A second order polynomial was fitted as an estimate of the mean CV over the range of repeat lengths challenged.

based on the error rate associated with the repeat length measurements, because clinical interpretations were less standardized during the earlier time period. Between 2003 and 2007, 1,511 assessments were performed, with 27 errors identified (1.8%, 95% CI: 1.2–2.6%). This is not significantly different from the 2.6% overall rate observed in the later time period (exact two tailed, $P = 0.3$). Nine errors were within one repeat of the acceptable range, whereas 10 were discrepant by five or more (0.7% vs. 5% in the later time period, $P = 0.4$). The proportion of participants with any error was slightly but not significantly higher (26% vs. 8% in the later time period, $P = 0.8$).

Precision profile for repeat length for US participants

The current grading scheme for repeat length is based on expert opinion.[3] Given the experience of this survey, it should now be possible to provide a more evidence-based approach to a grading that takes into account observed performance. **Figure 3** displays a precision profile for reported CAG repeat lengths for the 33 US participants after applying a two-pass, three standard deviation trimming algorithm. The pattern is "U" shaped, with highest precision at repeat lengths near the important cutoff level indicating the diagnosis of HD (40 repeats). Somewhat lower precision is found for smaller and larger repeat lengths. A set of integer repeat length ranges to include approximately 95% of the observations represented in this precision profile would be approximately ±2 repeats through 50 repeats and ±3 repeats for lengths of 51–75 repeats. When possible, repeat level cutoffs were chosen to be consistent with existing ACMG guidelines. Although no challenges occurred above 75 repeats, it seems reasonable that a less restrictive precision profile (e.g., ±4 repeats) would be a reasonable extrapolation. These proposed criteria are less restrictive than existing criteria promulgated by the ACMG and by the European Molecular Genetics Quality Network

Table 2 Total error rates for participants in the United States, stratified by primary analytic method

Analytic technique(s)	No.			Error rate (%)
	Participants	Errors	Assessments	
PCR, with capillary gel electrophoresis	26	17	814	3.0
PCR, with gel electrophoresis	6	2	210	1.1
PCR, bisulfite treatment	1	9	36	25.0
Any	33	28	1,060	2.6

Table 3 Comparisons of existing and proposed guidelines for assessing clinical laboratory performance of quantifying the Huntington disease CAG repeat length

ACMG		EMQN		Proposed	
CAG repeats	Limits	CAG repeats	Limits	CAG repeats	Limits
≤43	±1	≤42	±1		
44–50	±2	>42	±2	≤50	±2
51–75	±3			51–75	±3
>75	±4			>75	±4

ACMG, American College of Medical Genetics; EMQN, European Molecular Genetics Quality Network.

(EMQN, Table 2). Use of these proposed criteria would reduce the error rate in the US participants from 2.6 to 1.5%. Minor errors occurring in the normal range would not be identified, whereas all errors outside of the acceptable range by 2 or more repeats would remain classified as errors.

DISCUSSION

This is the first time that a comprehensive examination of the analytic validity of molecular testing for HD has been performed. It uses proficiency testing data collected by CAP/ACMG over an 8-year time period and focuses primarily on the performance of US participants. Such data provide documentation of the reliability of genetic testing in high complexity, certified laboratories that participate in external proficiency testing. These findings are likely to have included a high proportion of all laboratories offering clinical testing for HD in the United States, although this is difficult to formally confirm. No commercial kits/reagents are currently available in the United States because of the low volume of testing. Thus, all the results considered in this study have been derived from so-called "laboratory developed" tests. Three other external proficiency testing surveys for HD testing were identified: the UK National External Quality Assessment Scheme,[10] EMQN,[11] and a scheme administered by the Royal College of Pathology of Australasia.[12] Each of these programs offers three challenges once a year, compared with the three challenges twice a year offered by the CAP/ACMG proficiency testing survey.

Basing estimates of analytic validity on proficiency testing results has both strengths and weaknesses. The advantages lie in a reasonably comprehensive survey of laboratories that perform HD testing, with the corresponding wide representation of test methodologies that are in current use. Results from external proficiency testing surveys also allow estimates of pre- and postanalytic errors. These characteristics are difficult to obtain from other sources, including the published literature. There are, however, some differences between routine clinical laboratory testing and proficiency sample testing that could impact performance. Proficiency testing samples may be handled differently from clinical samples. For example, clinical samples are likely to be entered into an automated system that generates and sends reports without the need for hand transcription. Results of proficiency sample testing must

be hand entered in online report forms, using codes provided by CAP that provide the potential for postanalytic errors that would not happen with clinical samples. Such surveys are educational in nature, and, occasionally, less common genotypes may be overrepresented. For example, HD samples in the range of 26–39 repeats are far less common than the normal (or abnormal) sized alleles in the usual clinical flow of samples. However, they may be targeted for distribution to verify that laboratories can provide appropriate analytic interpretations in such circumstances.

The proficiency testing data collected by CAP/ACMG and their subsequent analysis shows very high analytic performance for participants located within the United States. Based on the participants' HD interpretation, analytic sensitivity and specificity are estimated to be 99.5% and 99.2%, respectively. The few errors seem not to be due to pre- or postanalytic errors but are more likely to be due to methodology (allelic dropout and sizing errors). An analysis of allele-specific repeat length finds an error rate of 2.6%, but the majority of these are in the normal range and outside of the acceptable range by only one repeat. These error rates have been consistent over time and are not dependent on the numbers of tests performed per month. However, one participant was responsible for nearly one third of all sizing errors from US participants. Were that one removed, the error rate for sizing would drop to 1.9%. We recommend that any laboratory using bisulfate PCR methodology for HD sizing either adopt an alternative, more standard methodology or cease testing. Starting in 2011, the HD survey allele sizes will also be graded according to the revised criteria provided in Table 3.

Nearly all US participants return results within 3 weeks. This is an acceptable timeframe for HD testing in the clinical settings of diagnosis and prognosis; the main indications for testing. This may not, however, be an acceptable timeframe for preconception and/or prenatal testing in an affected family. Checking with the laboratory on specific turn-around times may be warranted, as several participants report turn-around times of 7 days or less. Overall, the US participants report performing approximately 5,000 tests per year (423/month over 12 months). Assuming the prevalence of HD to be 3/100,000, approximately 9,300 individuals in the US

ORIGINAL RESEARCH ARTICLE

population of 310 million may be affected. This rate of testing provides evidence that HD testing is being used appropriately in differential diagnosis and not being used as a routine screening test in pregnancy; a clinical scenario that is not appropriate.

A further examination of the data suggests that the currently proposed criteria for laboratory precision in sizing repeats may be too stringent,[3] especially for repeat lengths <25. It may not be possible, or necessary, to require participants to be within one repeat at this level, and we suggest that this criterion be relaxed, while maintaining the stringent criteria for larger repeat lengths. This is especially relevant around 40 repeats, an important point for decision making. Were this new grading criterion implemented, the sizing error rate for US participants would be further reduced to <1% (after exclusion of the one poorly performing laboratory). Although this laboratory had difficulty in accurately sizing repeat lengths, their grading based on the analytic interpretation was acceptable. The proposed precision required by the criteria proposed by the EMQN[13] would be difficult to achieve, and there is little evidence it is necessary.

Analytic performance of the international participants is less good. Turn-around time was greater than 3 weeks for nearly half of the participants. A much higher number of sizing errors occurred (7.9%), many of which would lead to errors in interpretation. It is not clear why these participants had poorer analytic performance. This finding was also not consistent with the high analytic performance for HD testing found in the 2009 European Molecular Genetics Quality Network summary report. Among the 90 participants and 270 challenges, there were three "serious genotyping errors" (1.1%, 95% CI: 0.2–3.2%). Given the general availability of specific HD reference material for sizing,[14] it is unclear why these international laboratories participating in the CAP/ACMG survey are not performing to a higher standard. Information was not collected to determine whether these international participants were licensed by CLIA, and certified by CAP, New York State Department of Health, or an equivalent governmental/professional organization. Regardless, it is clear that proper interpretation of results from CAP/ACMG surveys may need to stratify results by geographic location because of the differences seen in the current HD survey analysis. In conclusion, US laboratories are performing at a high level in HD testing. Although not focused on HD testing per se, these findings are in sharp contrast to a 2006 report[4,5] that raised concerns about the quality of genetic testing in US laboratories.

DISCLOSURE

Dr. Richards' laboratory performs Huntington disease testing and participates in the CAP/ACMG proficiency testing survey. No other conflicts exist.

REFERENCES

1. Hayden M. Basal ganglia disorders. In: Rimoin DL, Connor JM, Pyeritz RE, Korf BR, (eds). *Emery and Rimoin's Principles and Practice of Medical Genetics*. Churchill-Livingstone: New York, 1996:2203–2208.
2. GeneTests. http://www.ncbi.nlm.nih.gov/sites/GeneTests. Accessed 28 March 2011.
3. American College of Medical Genetics, Standards and Guidelines for Clinical Genetics Laboratories; 2006 Edition. Technical standards and guidelines for Huntington disease. http://www.acmg.net/Pages/ACMG_Activities/stds-2002/HD.htm. Accessed 28 March 2011.
4. Hudson KL. Genetic testing oversight. *Science* 2006;313:1853.
5. Hudson KL, Murphy JA, Kaufman DJ, Javitt GH, Katsanis SH, Scott J. Oversight of US genetic testing laboratories. *Nat Biotechnol* 2006;24:1083–1090.
6. Robins Wahlin TB. To know or not to know: a review of behaviour and suicidal ideation in preclinical Huntington's disease. *Patient Educ Couns* 2007;65: 279–287.
7. Potter NT, Spector EB, Prior TW. Technical standards and guidelines for Huntington disease testing. *Genet Med* 2004;6:61–65.
8. Ehrich M, Zoll S, Sur S, van den Boom D. A new method for accurate assessment of DNA quality after bisulfite treatment. *Nucleic Acids Res* 2007;35:e29.
9. Grunau C, Clark SJ, Rosenthal A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res* 2001;29:E65–E65.
10. United Kingdom National External Quality Assessment Scheme (NEQAS). http://www.ukneqas-molgen.org.uk/ukneqas/index/news.html. Accessed 28 March 2011.
11. European Molecular Genetics Quality Network (EMQN). http://www.emqn.org/emqn/Schemes. Accessed 28 March 2011.
12. Royal College of Pathologists of Australasia (RCPA). http://www. rcpaqap.com.au/gendocs/uploadedfiles/InformationBooklet.pdf. Accessed 28 March 2011.
13. Best Practice European Molecular Genetics Quality Network (EMQN). Draft available at http://www.cmgs.org/BPGs/pdfs%20current%20bpgs/HD%20BPG%20April%202010%20for%20consultation.pdf. Accessed 28 March 2011.
14. Kalman L, Johnson MA, Beck J, et al. Development of genomic reference materials for Huntington disease genetic testing. *Genet Med* 2007;9:719–723.