ANALYTIC VALIDITY

Question 8: Is the test qualitative or quantitative?
Question 9: How often is a test positive when a mutation is present?
Question 10: How often is the test negative when a mutation is not present?
Question 11: Is an internal QC program defined and externally monitored?
Question 12: Have repeated measurements been made on specimens?
Question 13. What is the within- and between-laboratory precision?
Question 14: If appropriate, how is confirmatory testing performed to resolve false positives in a timely manner?
Question 15: What range of patient specimens has been tested?
Question 16: How often does the test fail to give a useable result?
Question 17: How similar are results obtained in multiple laboratories using the same, or different, technology?

ANALYTIC VALIDITY

Question 8: Is the test qualitative or quantitative?

In prenatal screening for cystic fibrosis, the aim is to identify couples in which both the mother and her partner have identifiable cystic fibrosis mutations. Their offspring have a 1 in 4 risk of having cystic fibrosis and definitive diagnostic testing is available. The DNA test results are qualitative (e.g., a specific mutation is reported as present or absent).

ANALYTIC VALIDITY

Question 9: How often is a test positive when a mutation is present?
Question 10: How often is the test negative when a mutation is not present?

Summary

External proficiency testing schemes are the only major reliable source currently available for computing analytic sensitivity and specificity. The following caveats should be kept in mind, however, when examining these estimates. First, external proficiency testing schemes are designed to be educational. For that reason, 'difficult' samples are over-represented. Also, laboratories from outside the U.S. are included, and both research and clinical laboratories participate. In spite of these shortcomings, this source of data can be useful in establishing a baseline of performance for laboratories.

Based on data from the American College of Medical Genetics and the College of American Pathologists (ACMG/CAP) Molecular Genetics Survey Set MGL

- The analytic sensitivity is 97.9% (95 percent CI 96.9 to 98.7%), after removing challenges involving delI507
- The analytic specificity is 99.4% (95 percent CI 98.7 to 99.8%), after removing challenges involving delI507 and adjusting for the rate of wrong mutations
- The analytic sensitivity and specificity are essential constant between 1996 and 2001

Based on data collected by the European Concerted Action on Cystic Fibrosis

- The overall raw error rate is 2.8% (95 percent CI 2.4 to 3.4%), consistent with raw error rates in the ACMG/CAP MGL Survey (3.0%, 95 percent CI 2.4 to 3.9%)
- Although all errors were reported (raw error rate), the reports did not distinguish between the type of error (e.g., false negative or false positive). For that reason, analytic sensitivity and specificity could not be determined
- Over three years of the program, performance steadily improved
- Over three years of the program, only 48 percent of laboratories made no errors
- Most (about 90 percent) of the errors occurred during the analytic phase of testing

Definitions

Analytic performance is summarized by the sensitivity and specificity of the detection system. Analytic sensitivity is the proportion of positive test results, when a detectable mutation is present (i.e., the test is designed to detect that specific mutation). The analytic sensitivity may also be called the analytic detection rate. Another way of expressing analytic sensitivity would be the true positives divided by the sum of the true positives and false negatives. False negative results could be due to technical errors in the analytic phase (e.g., sample placement, contamination, expired reagents and cross-reactivity) or to administrative/clerical errors in the pre-analytic or post-analytic phases (e.g., incorrect interpretation of correct analytic result, sample mislabeling and incorrectly copying a correct result).

Analytic specificity is the proportion of negative test results when no detectable mutation is present. Analytic specificity can also be expressed in terms of the analytic false positive rate. This would be the proportion of positive test results when no detectable mutations are present (1-

analytic specificity). Another way of expressing analytic specificity would be the true negatives divided by the sum of the true negatives and false positives. False positive results could be due to technical errors in the analytic phase (e.g., errors in placement, contamination, expired reagents, or non-specific reactions) or to administrative/clerical errors in the pre-analytic or post-analytic phases (e.g., mislabeling of samples, wrong interpretation of correct results, or copying results incorrectly).

Wrong mutations are a third type of error, along with false negative and false positive results. These occur when a mutation is present, but is incorrectly identified. For purposes of this review, wrong mutations will be considered false positive results, since there is an opportunity for correcting them by confirmatory testing. Wrong mutations occurring in the pre-analytic, analytic or post-analytic phases are all included in the analysis.

Optimal source(s) of data

Few data sources exist for estimating analytic validity. Published reports of method comparisons and screening experiences provide limited information on only a few testing methodologies. The data are derived from a small number of laboratories and the "true" genotypes of the tested samples are often uncertain (e.g., not confirmed by another methodology, laboratory consensus or sequencing). External proficiency testing programs (e.g., ACMG/CAP molecular Surveys and the European Concerted Action for Cystic Fibrosis) provide a source of data that have several advantages. They include a large proportion of clinical testing laboratories that represent the range of methodologies presently being used. In addition, the samples distributed have confirmed genotypes. However, basing analytic performance estimates on external proficiency testing also has drawbacks, including:

- over-representation of 'difficult' samples, due to the educational nature of the proficiency testing program
- mixing of screening and diagnostic exercises
- few challenges which do not contain a detectable mutation
- reporting summary results in ways that do not allow a straightforward computation of analytic sensitivity and specificity
- an important proportion of laboratories participating in the ACMG/CAP program are from outside the United States
- artificial nature of sample preparation, shipping and handling to ensure stability
- some participating laboratories involved with research or manufacturing rather than clinical activities

One additional consideration might be that laboratories perform differently when testing proficiency testing samples than when testing clinical samples on a routine basis. This difference might take the form of less good performance because the sample is handled outside of the laboratory routine. Alternatively, the performance might be better because extra attention might be paid to obtaining a reliable result. Future analyses should be aimed at providing reliable method- and, possibly, mutation-specific analytic performance estimates. One approach for collecting such data might include the following steps:

- An independent body would develop a standard set of samples, most of which would be randomly selected from the general population. Included in the standard set, however, would also be additional, less common genotypes (e.g., rarer heterozygotes, homozygotes and

compound heterozygotes). Sub-cloned samples are inadequate for this use. The group that collects and administers these samples and the subsequent analyses could be under the auspices of the FDA, ACMG, or CAP, or be a non-profit institution such as Coriell Institute for Medical Research (Camden, NJ). This effort would need grant support to begin the process.

- The sample set would then be available for method validation. Correct genotypes would be arrived at by consensus, or, if disagreements emerged, by a reference method (e.g., sequencing). The current validation practice of having a laboratory (or manufacturer) run a series of samples with unknown genotype (as is often the case for computing specificity) is inadequate, because there has been no comparison to a 'gold standard' (e.g., sequencing). For example, how can a laboratory running an unknown sample determine whether a positive finding is a true, or a false, positive or, whether a negative finding is a true, or false, negative?
- Ideally, this blinded sample set would be available to manufacturers as part of the pre-market approval process, with the understanding that multiple laboratories using these commercial reagents would be asked by the manufacturer to analyze portions of the sample set independently. This initial assay validation process is distinct from assay control samples that are discussed later (Question 11).

Appropriate sample size for determining analytic specificity can be derived by choosing an acceptable target specificity and an acceptable lower limit that should be excluded in the 95 percent confidence interval. The higher the specificity chosen and the tighter the confidence interval, the larger is the sample size that will be necessary to provide a definitive answer. For example, if a laboratory chose a target specificity of 98 percent and wanted to rule out a specificity of 90 percent, it would need to correctly identify at least 49 of 50 known negative samples (estimated using the binomial distribution). On the other hand, a target specificity of 99.5 percent and a desire to rule out a specificity of 98 percent would require correctly identifying at least 398 of 400 known negative samples. The determination of even higher analytic specificity with tighter confidence intervals may not be economically feasible for an individual laboratory. However, this could be attained by a consortium of laboratories using the same methodology, or by a manufacturer that forms a consortium of laboratories using its reagents.

Appropriate sample size for determining the analytic sensitivity (detection rate) could be derived using similar analyses. If a laboratory chose a target sensitivity of 95 percent and wanted to rule out a sensitivity of 80 percent, it would need to correctly identify at least 38 of 40 chromosomes with known mutations. A higher sensitivity estimate of 98 percent that rules out a rate of 95 percent would require the correct identification of at least 196 of 200 chromosomes with known mutations. If mutation-specific detection rates are desired, each would need the same number of challenges. Again, however, this may not be feasible for individual laboratories but may be possible for a consortium or manufacturer, especially for the more common mutations.

The analytic performance (analytic sensitivity and specificity) could then be determined for each methodology, along with an estimate of between-laboratory, within-method variability. Further, estimates could be made for specific racial/ethnic groups, based on the mutation-specific performance and the frequency of each mutation within that group. Overall, the analytic

performance for laboratories in the United States could be estimated, given the mix of methodologies for established screening laboratories. All of these analyses could be done using a 2x2 table, and all rates could be accompanied by 95 percent confidence intervals (CI). Published method comparisons focus on technical errors in the analytic phase and usually do not deal with the pre- and post-analytic phases of the laboratory testing process.

The ACMG/CAP external proficiency testing scheme

Background and definitions As part of ACMG/CAP external proficiency testing in the United States, purified DNA from established cell lines (derived from human cells with known mutations http::/locus.umdmj.edu/nigms/qc/dnaqc.html) is distributed to enrolled laboratories. The majority of these laboratories are likely to be providing clinical services, but reagent manufacturers and research laboratories also participate. In late 2001, there were 45 participants reporting cystic fibrosis results. A false positive result occurs when the laboratory reports finding a mutation in the sample, when none is present. A false negative result occurs when a laboratory reports no mutation, but a mutation for which it tests is, in fact, present in the sample. A third type of error occurs when the laboratory accurately identifies that a mutation is present, but it is not the correct mutation (e.g., a laboratory that is able to separately identify delF508 and delI507 reports finding delF508 when only the delI507 mutation is present). The three types of errors all are included in the analysis and encompass all three phases of testing.

The present analysis, which utilizes the ACMG/CAP data, initially examines the rates of these three types of errors independently, by chromosome (e.g., the results on one chromosome are counted separately from the results reported for the other).

> **Gap in Knowledge: How should the finding of a wrong mutation influence computation of the analytic performance?** The relationship between the third type of error (wrong mutation) and analytic performance has not yet been formally addressed. In this document, a wrong mutation will be considered an incorrect result, since this type of error could cause harm. For example, diagnostic testing in the fetus might target the mutations reported in the couple and not identify the correct mutation in the fetus. Also, family members would not receive correct information. Further, a wrong mutation finding will treated as a false positive in this document. Confirmatory testing of positive results will provide the opportunity to correct this type of error

Error rates for the ACMG/CAP external proficiency testing scheme Table 2-1 shows the number of alleles tested and the results from the ACMG/CAP Molecular Genetics Laboratory (MGL) Survey from 1996 to 2001. Overall, 3.0% (95 percent CI 2.4 to 3.9%) of the alleles were incorrectly identified. For all data between 1996 and 2001, 2,131 of 2,198 chromosomes 97.0 percent were correctly identified (95 percent CI 96.1 to 97.6%). Appendix A contains a complete listing of the sample challenges, the responses along with the type of error (e.g., false positive), and any other adjustments made during the analysis (e.g., laboratory did not test for a mutation included in the challenge). More errors (56) occurred between 1996 and 1998 than between 1999 and 2001 (11). However, the composition of challenges in the earlier time period explains much of this excess and is taken into account in analyses that are presented later in this section.

Table 2-1. *CFTR* Mutation Testing: Results of the ACMG/CAP MGL Survey

Year	Number Of Labs	Alleles Tested	Correct N (%)	Incorrect N (%)	Type of Incorrect Result: False Positive N (%)	False Negative N (%)	Wrong Mutation N (%)
1996	47	282	267 (96.5)	15 (3.5)	2 (0.7)	2 (0.7)	11 (3.9)
1997	46	276	245 (89.5)	31 (10.5)	6 (2.2)	7 (2.5)	18 (6.5)
1998	51	306	296 (96.7)	10 (3.3)	0 (0.0)	10 (3.3)	0 (0.0)
1999	43	342	341 (99.7)	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.3)
2000	41	458	452 (98.7)	6 (1.3)	0 (0.0)	2 (0.4)	4 (0.9)
2001	45	534	528 (99.2)	4 (0.8)	2 (0.4)	1 (0.2)	1 (0.2)
All		2198	2131 (97.0)	67 (3.0)	10 (0.5)	22 (1.0)	34 (1.6)

Table 2-2 makes use of the ACMG/CAP MGL Survey data (Appendix A) to compute a preliminary estimate of analytic sensitivity and specificity. The apparent improvement in performance over time may be real, or due to differences in the types of challenges. For example, no wild/wild mutation challenges were included prior to 2000, while 8 of 12 challenges since then were wild/wild. It is not possible, because of the small numbers, to stratify the results by methodology or to provide separate estimates of performance for most of the mutations tested.

Table 2-2. Analytic Performance for Identifying All Cystic Fibrosis Mutations According to Data from the ACMG/CAP Molecular Genetics Survey

Year	Analytic Sensitivity (%)	(95% CI)	Analytic Specificity	(95% CI)
1996	98.9	(96.1-99.9)	87.1	(79.0-93.0)
1997	96.0	(91.8-98.4)	76.7	(67.3-84.5)
1998	96.5	(93.6-98.3)	100.0	(83.9- 100)
1999	100.0	(98.3- 100)	99.2	(95.8-99.9)
2000	97.4	(90.8-99.7)	99.0	(97.3-99.7)
2001	99.4	(96.7-99.9)	99.2	(97.6-99.8)
All	97.9	(96.9-98.7)	95.8	(94.4-96.9)

Complicating factors in interpreting these results An additional aim of these external challenges was education. For that reason, it may not be appropriate to use these data to determine analytic performance without taking into account the design of these exercises. For example, 14 percent (3/21) of the challenges required that participating laboratories distinguish between the delI507 and delF508 mutations. All of these challenges occurred in the first three years of the survey. The delI507 mutation occurs in less than 1 in 2500 non-Hispanic Caucasians tested (1 percent of 1/25). This rare and difficult laboratory circumstance is emphasized because of the educational and laboratory-improvement focus of the ACMG/CAP MGL Survey. An additional complicating feature arises because it is not always clear whether some 'false negatives' might

be due to laboratories not testing for the mutation. The present analysis attempts to take this into account (Appendix A). The opportunity for a laboratory to identify a wrong mutation is considerably greater in proficiency testing exercises than in practice, due to the high frequency of mutations. For that reason, the rate of wrong mutations in proficiency testing needs to be adjusted downward in order to simulate performance in routine clinical practice.

A more reliable approach to estimating analytic sensitivity and specificity It is possible to recompute the previous analysis using only challenges that do not involve delI507. Separate estimates can then computed for the four challenges involving delI507. These two stratified estimates of analytic performance are shown in Table 2-3, along with the summary estimate from Table 2-2. The analytic specificity for identifying the delI507 mutation is poorer than for the other mutations. The sensitivity is actually better, since some mutation was reported in all instances where a delI507 mutation was present. A better estimate of overall performance that would be expected in the real world is found when challenges involving the delI507 mutation are not counted (the bolded row in Table 2-3).

Table 2-3. Analytic Performance With and Without delI507 Mutation Challenges Based on the ACMG/CAP Molecular Genetics Survey Data

Mutation Group	Challenges	Sensitivity (%)	Specificity (%)
All mutations (Table 2-2)	2198	97.9 (96.9-98.7)[1]	95.8 (94.4-96.9)
All but delI507	1940	**97.9 (96.8-98.7)**	98.4 (97.3-99.1)[2]
delI507 only	258	100 (97.1-100)	79.1 (71.2-85.6)

[1] 95 percent CI
[2] A more reliable estimate of analytic specificity is provided later in this section.

Table 2-4 shows the analytic performance estimates by year for challenges without delI507. No trend is evident for improvement in analytic sensitivity, and the overall rate of 97.9 percent appears reasonable. The upper and lower confidence intervals could be taken to model the most pessimistic (96.8 percent) and optimistic (98.7 percent) estimates of analytic sensitivity. A standardized mutation panel is now becoming widely adopted, as a result of ACMG recommendations (Grody WW, 2001). As a result, manufacturers are now marketing reagents (under the rule for Analyte Specific Reagents – ASR) that have been subjected to good manufacturing processes. Analytic performance may improve as a consequence. The present analysis establishes a 'baseline' estimate of analytic sensitivity and specificity, against which to assess that possibility.

Analytic specificity is more difficult to interpret. Thirteen of 15 errors occurred during one distribution (1997-B). Some of these might be explained by sample mix-up, but at least half appear not to be due to this cause. The European Concerted Action on Cystic Fibrosis reported that commercial kits were found to have problems identifying G551D and R553X. The majority of errors in the 1997 ACMG/CAP survey occurred when challenging these two mutations.

Table 2-4. Analytic Performance for Cystic Fibrosis Mutations According to Data from the ACMG/CAP Molecular Genetics Survey (Excluding delI507 Mutation Challenges)

Year	Analytic Sensitivity (%)	(95% CI)	Analytic Specificity (%)	(95% CI)
1996	98.5	(94.8-99.8)	98.1	(89.9-99.9)
1997	96.1	(91.1-98.7)	82.5	(70.1-91.3)
1998	96.5	(93.6-98.3)	100	(83.9-100)
1999	100	(98.3-100)	99.2	(95.8-99.9)
2000	95.3	(84.2-99.4)	100	(98.9-100)
2001	99.4	(96.7-99.9)	99.2	(97.6-99.8)
All but delI507 (Table 2-3)	**97.9**	**(96.8-98.7)**	98.4[1]	(97.3-99.1)

[1] A more reliable estimate of analytic specificity is provided later in the next section

A final estimate for analytic specificity As stated earlier, the definition being used in this analysis for false positives (1-specificity) are composed of two types of errors: false positive results and wrong mutations. Finding a 'false positive' can occur whenever a detectable mutation is not present; a common situation in screening. The finding of a 'wrong mutation' can only occur when a mutation is present; a relatively uncommon common situation in screening. However, it is common in proficiency testing samples. There have been a total of 949 mutation challenges and 922 wild challenges (after ignoring all delI507 samples). Thus, a mutation being tested for is present in about 50 percent of the chromosomes. Conversely, only about 1.8 percent of chromosomes in the general pregnancy population will have a mutation identified (1/25 non-Hispanic Caucasians are carriers and about 90 percent of the mutations on the mutated chromosome can be detected). For this reason, the rate of wrong mutations must be 'discounted' by a factor of about 28 (50/1.8). Thus, although Table 2-1 shows a ratio of 10 false positive results to 34 wrong mutations, the expected ratio in the general population would be more like 10 false positives to 1 or 2 wrong mutations (34/28). After samples have been removed that included delI507 and after the rate of 'wrong mutation' in the general population has been taken into account, the revised estimate of analytic specificity is 99.4% (95 percent CI 98.7 to 99.8%).

Gap in Knowledge: Method- and mutation-specific analytic performance estimates
Tables 2-2 through 2-4 present the best available data for estimating analytic performance. These analyses should not be interpreted as being complete or robust. For example, the problems identified by the delI507/delF508 challenges are method-specific, but no attempt is made in this report to analyze laboratory performance by specific method. The results here are for the mix of methodologies presently being used in the United States and, as such, represent the average laboratory performance a clinician might expect when ordering such testing. To generate more reliable analytic performance estimates, large numbers of specimens with known genotypes will need to be run using specific methodologies. For example, Gasparini *et al*. (1999) used the PCR/OLA methodology to identify 114 newborns with a mutation; all of these were subsequently confirmed by DNA sequencing. Although this rules out false positives, it does not provide an estimate of analytic sensitivity, since only a small random subset of negative results was similarly sequenced and the possibility of false negative results

exists. Until more refined performance estimates are available, the existing information is useful in estimating clinical performance.

Gap in Knowledge: Analytic performance estimates are available for only a small number of mutations. Only a small number of mutations (10) has been subjected to external proficiency testing (delF508, delI507, G542X, 621+1G>T, G85E, W1282X, G551D, R553X, 1717-1G>T, and R117H). The majority of the mutations in the recommended panel have not have been subjected to external proficiency testing. This is an important consideration because performance may vary according to laboratory methodology.

Gap in Knowledge: Analytic performance and mutation panel size. It is possible that analytic performance will differ, depending on the numbers of mutations tested, even when the same methodology is employed. Panels utilizing a higher number of mutations might be more robust because of automation or, conversely, the larger number of analytic steps might be more prone to errors.

Sensitivity and specificity by person rather than by chromosome

It is possible to compute analytic sensitivity and specificity according to whether a person's genotype has been correctly classified, rather than whether an individual chromosome has been correctly classified. That is, the genotype is correct or incorrect when detectable mutations are present (analytic sensitivity) or the genotype is correct or incorrect when no detectable mutations are present (analytic specificity). Table 2-5 shows the results of this analytic approach, stratified by the year that proficiency testing results were obtained. All three samples containing a delI507 mutation have been removed from the analysis. According to these data, the overall estimate for analytic sensitivity is 95.9% (95 percent CI 93.3 to 97.1%). This is lower than shown in Table 2-4 (97.9 percent), where the analysis is by chromosome rather than by person. When the analysis is performed by person, wrong mutations are included in the computation of analytic sensitivity. Once the eight instances of wrong mutations are accounted for, analytic sensitivity is corrected upward to 97.2 percent. This estimate is now similar to that found when the analysis was by chromosome. Table 2-5 also shows an analytic specificity of 99.7% (95 percent CI 98.4 to 99.9%), consistent with that found in Table 2-4 (99.4 percent).

Table 2-5. Analytic Sensitivity and Specificity based on the ACMG/CAP MGL Survey, Classified According to Whether a Person's Genotype is Correctly Identified

	Correct N (%)	Incorrect N (%)	Totals
Detectable mutation present			
1996	91 (96.8)	3 (3.2)	94
1997	83 (90.2)	9 (9.8)	92
1998	143 (93.5)	10 (6.5)	153
1999	171 (99.4)	1 (0.6)	172
2000	32 (97.0)	1 (3.0)	33
2001	87 (97.8)	2 (2.2)	89
Analytic Sensitivity	**607 (95.9)**	**26 (4.1)**	**633**
Detectable mutation not present			
1996	3 (100)	0 (0.0)	3
1997	9 (100)	0 (0.0)	9
1998	1 (100)	0 (0.0)	1
1999	2 (100)	0 (0.0)	2
2000	155 (100)	0 (0.0)	156
2001	171 (99.4)	1 (0.6)	172
Analytic Specificity	**341 (99.7)**	**1 (0.3)**	**342**

External proficiency testing in Europe

Results of the proficiency testing survey conducted by the European Concerted Action for Cystic Fibrosis. Table 2-6 shows the results of that study. Because that study's report did not distinguish between false positive, false negative and incorrect mutations, it is not possible to compute an analytic sensitivity or specificity. However, the overall rate of 2.8 percent incorrectly classified chromosomes (95 percent CI 2.4 to 3.4%) is similar to the overall 3.0 percent error rate found in the ACMG/CAP survey reported earlier in this section. This study also reported that 48 percent of 114 participants had correct responses for all challenges. Another 39 percent committed one error, while 2 percent failed all challenges.

Interpretation of the results. This survey also attempted to determine the cause of errors, including sample contamination and clerical errors. In general, laboratories would have been able to correct their false positive results, if their policy had been to reanalyze samples with positive results. This indicates that the original sample was neither contaminated nor incorrectly labeled. Clerical errors/reporting mistakes/incorrect interpretations were estimated to be responsible for 90 percent of the errors. The error rate was not associated with the numbers of samples processed by the laboratory.

Table 2-6. Survey Results from the European Concerted Action for Cystic Fibrosis, According to Whether the Chromosome was Correctly Classified

Year	Alleles Tested	Correct N (%)	Incorrect N (%)
1996	1632	1569 (96.1)	63 (3.9)
1997	1740	1691 (97.2)	49 (2.8)
1998	1908	1872 (98.1)	36 (1.9)
All	5280	5132 (97.2)	148 (2.8)

Comparing error rates for DNA-based cystic fibrosis testing with biochemical testing for Down syndrome

A similar proficiency testing program (Survey FP) for maternal serum Down syndrome markers serves as one source for comparing error rates in non-DNA testing. In that survey (jointly sponsored by the Foundation for Blood Research and CAP), participating laboratories are asked to measure three biochemical markers, to combine these measurements with a pre-assigned maternal age, and then calculate a Down syndrome risk. Five challenges are distributed, three times each year. The proportion of laboratories with one or more outlying Down syndrome risk estimates on a given distribution is routinely reported to all participants each year (FBR/CAP FP Survey Participant Summary Report, 2000, FP-C). This proportion has remained relatively constant between 1998 and 2000 at about 5 percent. Assuming that the laboratory will have only one (or two) of the five risks classified as being an outlier, the actual error rate per sample distributed is closer to 1 or 2 percent. This is similar to the error rate for the ACMG/CAP MGL survey found in Table 2-1. This analysis is limited to data prior to 2001, since a problem with sample preparation was identified in 2001 and corrected in 2002.

References

ACMG/CAP Molecular Genetics Survey Sets. 1996, 1997, 1998, 1999, 2000, 2001. College of American Pathologists, Northfield, IL.

Cuppens H, Cassiman JJ. 1995. A quality control study of CFTR mutation screening in 40 different European laboratories. *Eur J Hum Genet* **3**:235-245.

Dequeker E, Cassiman J. 2000. Genetic Proficiency testing in diagnostic laboratories – quality control is the message. *Am J Hum Genet* **67**:A274.

Dequeker E, Cuppens H, Dodge J, Estivill X et al. 2000. Recommendations for quality improvement in genetic testing for cystic fibrosis European Concerted Action of Cystic Fibrosis. *Eur J Hum Genet* **8**:s2-s24.

FBR/CAP Maternal Serum Survey Sets (2000) College of American Pathologists, Northfield, IL.

Gasparini P, Arbustini E, Restagno, Zelante I, Stanziale P, Gatta L, Sbaiz L, Sedita AM, Banchieri N, Sapone L, Fiorucci GC, Brinson E, Shulse E, Rappaport E, Fortina P. 1999. Analysis of 31 CFTR mutations by polymerase chain reaction/oligonucleotide ligation assay in a pilot screening of 4476 newborns for cystic fibrosis. *J Med Screen* **6**:67-69.

Gordy WW, Cutting GR, Klinger KW et al. 2001. Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet Med* **3**:149-54.

Grody WW, Pyeritz RE. 1999. Report card on molecular genetic testing. *JAMA* **281**:845-7.

Grody WW, Desnick RJ, Carpenter NJ, Noll WW. 1998. Diversity of cystic fibrosis mutation-screening practices. *Am J Hum Genet* **62**:1254-8.

Hofgartner WT, Tait JF. Frequency of problems during clinical molecular-genetic testing. 1999. *Am J Clin Pathol* **112**:14-21.

Appendix A. Data used to calculate analytic sensitivity and specificity

Table 2-7 summarizes the cystic fibrosis external proficiency testing results obtained by the American College of Medical Genetics and the College of American Pathologists (ACMG/CAP). Samples with known genotypes have been distributed to participants since 1996. The first column of the table contains the distribution label (96 MGL-11 indicates the 11th DNA sample distributed as part of the Molecular Genetics Laboratory Survey in 1996). The second column contains number of participating laboratories, followed by the genotype of the sample. The number of laboratories reporting specific genotypes is then provided, along with a tabulation of their 'correct' and 'incorrect' responses. The last two columns provide an adjusted interpretation by taking into account that some laboratories do not test for all mutational challenges. The last column also shows the type of error: wrong mutation (wm - indicating that the laboratory reported a mutation, but not the correct one), false positive (fp – indicating a mutation was reported, when none was present) or false negative (fn – a tested mutation was present, but none was reported). In a few instances, it is possible that there was a sample mix-up (e.g., 1997 MGL-11/MGL-12 for 4 laboratories). However, no allowance for administrative error has been made in this analysis. The table also contains the analytic sensitivity and specificity in a box, along with the yearly (and summary) totals.

Table 2-7. Computations for the ACMG/CAP Proficiency Testing Surveys (continued)

			Reported Alleles		Adjusted Report	
Distribution	**Labs**	**Genotype**	**Correct**	**Incorrect**	**Correct**	**Incorrect**
96 MGL-11	**47**	**508/621**				
	42	508/621	84	0	84	0
	4	508/wild	4	4	7[1]	1 (fn)
	1	508/508	1	1	1	1 (fp)
96 MGL-12	**47**	**507/wild**				
	34	507/wild	68	0	68	0
	11	508/wild	11	11	11	11 (wm)
	1	508/507	1	1	1	1 (fp)
	1	wild/wild	1	1	2[2]	0
96 MGL-13	**47**	**542/wild**				
	45	542/wild	90	0	90	0
	2	wild/wild	2	2	3[2]	1 (fn)
Totals 1996		**282 alleles**	**262**	**20**	**267**	**15**
Sensitivity		84+5+1+34+11+1+0+45+0			179	2
Specificity		0+3+1+34+11+1+2+45+4			88	13

Table 2-7. Computations for the ACMG/CAP Proficiency Testing Surveys (continued)

Distribution	Labs	Genotype	Raw Alleles Correct	Raw Alleles Incorrect	Adjusted Alleles Correct	Adjusted Alleles Incorrect
97 MGL-10	**46**	**507/wild**				
	32	507/wild	64	0	64	0
	7	508/wild	7	7	7	7 (wm)
	3	wild/wild	3	3	4[2]	2 (fn)
	1	507/551	1	1	1	1 (fp)
	1	621/1303	0	2	0	2(1wm 1fp)
	1	1282/1303	0	2	0	2(1wm 1fp)
	1	560/621	0	2	0	2(1wm 1fp)
97 MGL-11	**46**	**551/553**				
	39	551/553	78	0	78	0
	4	1717/wild	0	8	0	8[3](4wm4fn)
	1	551/wild	1	1	1	1 (fn)
	1	1162/1282	0	2	0	2 (wm)
	1	wild/wild	0	2	2[2]	0
97 MGL-12	**46**	**1717/wild**				
	35	1717/wild	70	0	70	0
	8	wild/wild	8	8	16[4]	0
	1	551/117	0	2	0	2(1wm 1fp)
	1	1303/1717	1	1	1	1 (fp)
	1	553/wild	1	1	1	1 (wm)
Totals 1997		**276 alleles**	**234**	**42**	**245**	**31**
Sensitivity		32+7+2+1+1+1+1+78+8+2+2+0+35+ 0+1+2			166	7
Specificity		32+7+4+1+1+1+1+ 0+0+0+0+2+35+16+1+2			79	24

Table 2-7. Computations for the ACMG/CAP Proficiency Testing Surveys (continued)

Distribution	Labs	Genotype	Raw Alleles Correct	Raw Alleles Incorrect	Adjusted Alleles Correct	Adjusted Alleles Incorrect
98 MGL-04	**51**	**542/542**				
	42	542/542	84	0	84	0
	6	542/wild	6	6	6	6 (fn)
	2	542/-	2	2	4[5]	0
	1	wild/wild	0	2	2[2]	0
98 MGL-05	**51**	**508/621**				
	42	508/621	84	0	84	0
	8	508/wild	8	8	13[6]	3 (fn)
	1	621/wild	1	1	1	1 (fn)
98 MGL-06	**51**	**508/117**				
	39	508/117	78	0	78	0
	12	508/wild	12	12	24[7]	0
Totals 1998		**306 alleles**	**275**	**31**	**296**	**10**
Sensitivity		84+12+2+0+84+11+2+78+12			275	10
Specificity		0+ 0+2+2+ 0+ 5+0+ 0+12			21	0
99 MGL-03	**43**	**508/508**				
	43	508/508	86	0	86	0
99 MGL-04	**42**	**1282/wild**				
	40	1282/wild	80	0	80	0
	1	508/wild	1	1	1	1 (wm)
	1	wild/wild	1	1	2[2]	0
99 MGL-15	**43**	**1282/wild**				
	42	1282/wild	84	0	84	0
	1	wild/wild	1	1	2[2]	0
99 MGL-16	**43**	**508/wild**				
	42	508/wild	84	0	84	0
	1	508/-	1	1	2[8]	0
Totals 1999		**342 alleles**	**338**	**4**	**341**	**1**
Sensitivity		86+40+1+0+42+0+42+1			212	0
Specificity		0+40+1+2+42+2+42+1			129	1

Table 2-7. Computations for the ACMG/CAP Proficiency Testing Surveys (continued)

Distribution	Labs	Genotype	Raw Alleles Correct	Raw Alleles Incorrect	Adjusted Alleles Correct	Adjusted Alleles Incorrect
00 MGL-04	**35**	**wild/wild**				
	35	wild/wild	70	0	70	0
00 MGL-05	**36**	**507/wild**				
	29	507/wild	58	0	58	0
	3	wild/wild	3	3	6	0
	3	508/wild	3	3	39	3 (wm)
	1	508/-	0	2	1[9]	1 (wm)
00-MGL-06	**35**	**wild/wild**				
	35	wild/wild	70	0	70	0
00-MGL-16	**41**	**wild/wild**				
	41	wild/wild	82	0	82	0
00-MGL-17	**41**	**621/G85E**				
	20	621/G85E	40	0	40	0
	16	621/wild	16	16	32	0
	4	wild/wild	0	8	7	1 (fn)
	1	G85E/wild	1	1	1	1 (fn)
00-MGL-18	**41**	**wild/wild**				
	41	wild/wild	82	0	82	0
Totals 2000		**458 alleles**	**425**	**33**	**452**	**6**
Sensitivity		0+29+0+3+1+ 0+ 0+40+ 0+1+2+0			74	2
Specificity		70+29+6+3+1+70+82+ 0+32+7+0+82			378	4

Table 2-7. Computations for the ACMG/CAP Proficiency Testing Surveys (continued)

Distribution	Labs	Genotype	Raw Alleles Correct	Raw Alleles Incorrect	Adjusted Alleles Correct	Adjusted Alleles Incorrect
01 MGL-04	**44**	**wild/wild**				
	44	wild/wild	88	0	88	0
01 MGL-05	**44**	**508/R117H**				
	39	508/R117H	78	0	78	0
	5	508/wild	5	5	10[7]	0
01-MGL-06	**44**	**wild/wild**				
	44	wild/wild	88	0	88	0
01-MGL-16	**45**	**wild/wild**				
	45	wild/wild	90	0	90	0
01-MGL-17	**45**	**G551/R553**				
	40	G551/R553	80	0	80	0
	3	wild/wild	0	6	6[10]	0
	1	508/wild	0	2	0[11]	2 (1fn 1fp)
	1	G551/G551	1	1	1	1 (wm)
01-MGL-18	**45**	**wild/wild**				
	44	wild/wild	88	0	88	0
	1	508/wild	1	1	1	1 (fp)
Totals 2001		**534 alleles**	**519**	**15**	**530**	**4**
Sensitivity			0+78+5+0+0+80+2+2+0		167	1
Specificity			88+5+88+90+6+88+2		367	3
Totals 96-01		**2171 alleles**	**2021**	**120**	**2082**	**59**

[1] Three laboratories reported that they did not test for 621
[2] One laboratory tested only for 508
[3] Likely a sample mix-up, since the correct answer for MGL-12 was 1717 heterozygote
[4] Eight laboratories did not test for 1717
[5] Two laboratories did not test for the normal allele
[6] Only three of these laboratories tested for 621
[7] None of these laboratories tested for R117H
[8] Did not report the second allele – report format error
[9] According to ACMG/CAP report
[10] These laboratories did not test for either G551D or R553X
[11] Assumes that this laboratory did test for G551D and R553X

Many of the analyses in the preceding sections (Questions 9 and 10) were made possible by the data collection efforts of the ACMG/CAP MGL Committee who were responsible for all aspects of that proficiency testing program during the years 1996 through 2001. After reviewing Questions 9 and 10, the following response was received from the CAP/ACMG Biochemical and Molecular Genetics Resource Committee:

RESPONSE AND COMMENTARY OF THE CAP/ACMG BIOCHEMICAL AND MOLECULAR GENETICS RESOURCE COMMITTEE

June 23, 2002

Since its inception, the College of American Pathologists quality assurance program has had education as perhaps its primary goal. This is certainly true for proficiency testing surveys in the newer areas of laboratory medicine, such as molecular genetics, where the data accrued from participating laboratories represents the most comprehensive "snapshot" of the state of the art, technical trends and developments as these fields evolve and mature. The more use that can be made of the survey data, the greater the return on the considerable investment of money and effort that goes into designing and participating in these programs. CAP encourages its resource committees, as the authors of these proficiency surveys, to maximize return through creative data mining and participant feedback. The CAP/ACMG Biochemical and Molecular Genetics Resource Committee, source of the Molecular Genetics (MGL) survey series, has been pursuing this goal for several years, through data analysis and publishing, expanding the scope of methods-specific questionnaires sent with the survey materials, and tracking the source of errors in performance.

Thus, the act of analysis, synthesis, and re-analysis of participant performance data is encouraged. However, the Committee, which is most intimately familiar with the construction of these surveys, is also wary of the dangers of overinterpreting these data and overextrapolating laboratory performance based on limited proficiency testing experience to real-world accuracy of these tests in the field. The following is a summary of the Committee's response to some of the conclusions in the Analytic Validity section of the ACCE document on population-based cystic fibrosis screening which we have been invited to review:

1) The members of the CAP/ACMG Biochemical and Molecular Genetics Resource Committee appreciate the recognition that our molecular genetics proficiency testing experience represents the most systematic survey of laboratory performance now extant. On the other hand, we concur with the caveats pointed out by the authors that raw calculations of sensitivity and specificity based on these data may appear to impart more global validity to the conclusions about real-world performance than is justified based on our limited and somewhat artificial experience thus far.

2) It is well known that, despite our best efforts to achieve realism, the specimen types, method of delivery, and mode of workup and result reporting for proficiency test samples are usually not typical of a laboratory's standard workload, increasing the chances for spurious clerical and technical errors. Indeed, the Committee's preliminary investigation into the source of proficiency testing errors in some of our other surveys, such as factor V Leiden, revealed the majority to be of a clerical nature. While we realize that a wrong result from this cause is still a

wrong result, the raw CAP/ACMG data probably do not directly reflect the technical sensitivity/specificity of the assays themselves.

3) Much of the reported underdetection of CF mutations, especially in the early years of the survey, were actually due to labs not testing for the challenged mutation, rather than poor analytic sensitivity or technical errors. In fact, we addressed this by surveying the participating laboratories for the size and makeup of their CF testing panels (Grody W.W., R.J. Desnick, N.J. Carpenter and W.W. Noll. 1998. Diversity of cystic fibrosis mutation screening practices. *Am. J. Hum. Genet.* 62:1252-1254), and found that, prior to the recent recommendations for standardization of CF mutation screening panels, there was wide variance in numbers of mutations tested and prevalent use of what today would be considered substandard test panels.

4) The labs surveyed during the years used for the analysis were using a wide variety of home brew techniques and mutation panels, most of which have changed now that the ACMG 25-mutation panel became the official standard of care in late 2001. As pointed out in the document, several companies have begun to sell robust ASR products for detection of these mutations, which the vast majority of labs are now using. The analysis based on data from the years 1996-2001, therefore, may not reflect the present state of the art.

In summary, while the Committee appreciates the educational value of its participant survey data for assessing trends in the field and promoting laboratory quality, we are reluctant to extrapolate too much from performance of individual laboratories in proficiency testing programs to general laboratory performance in the field. This is especially true for a program as relatively young as the molecular genetics (MGL) survey and for a disease analyte (CFTR) that has so recently undergone a major change in its technical approach.

Respectfully submitted by the members of the
CAP/ACMG Biochemical and Molecular Genetics Resource Committee

ANALYTIC VALIDITY

Question 11: Is an internal QC program defined and externally monitored?

Summary
- Internal quality control procedures are well described in several published sources
- External monitoring is provided through inspections conducted by accrediting organizations such as CLIA, CAP or New York State
- There is an unmet need for positive control samples

Definition
Internal quality control is a set of laboratory procedures designed to ensure that the test method is working properly. An internal quality control program includes documentation that high standards are being practiced to ensure that:

- reagents used in all aspects of genetic testing are of high quality to allow successful test completion,
- all equipment is properly calibrated and maintained,
- good laboratory practices are being applied at every level of genetic testing. To the extent possible, all steps of the testing process must be controlled.

Quality control procedures
Techniques that are used for analyzing DNA in prenatal screening for cystic fibrosis are the same as those used for other molecular testing. These techniques are widely applied and well understood. As a result, it has been possible to design and publish generic internal quality control procedures, which many molecular laboratories already have in place. Table 2-8 lists published guidelines that, among other topics, describe reagent quality control, equipment calibration and maintenance, education of the technical staff, and other internal quality control procedures. The purpose of the quality control procedures is to rigorously control all steps of the DNA testing process to minimize the potential for test failure. Given that the internal procedures for establishing and maintaining good laboratory practice are readily available (Neumaier *et al*., 1998), the important next step will be to encourage, assist, and require laboratories to apply and document appropriate quality control procedures.

Table 2-8. Guidelines, Recommendations, and Checklists that Address Internal Quality Control Issues and Requirements.

Guidelines, Recommendations and Checklists	Source / Reference
Clinical Laboratory Improvement Amendments of 1988	Federal Register 1992;57:7002-3
Genetic Testing Under CLIA	Federal Register 2000;65: 25928-24934
New York State Laboratory Standards (9/00)	www.wadsworth.org/labcert/download.htm
Molecular Diagnostic Methods for Genetic Diseases: Approved Guidelines	National Committee for Clinical Laboratory Standards MM1-A Vol 20 #7
College of American Pathologists Checklist	www.cap.org
Standards and Guidelines for Clinical Genetics Testing	American College of Medical Genetics www.faseb.org/genetics/acmg/stds
European Concerted Action on Cystic Fibrosis	(BMH-4-CT96-0462)
Laboratory Standards and Guidelines for Population-based Cystic Fibrosis Carrier Screening	Grody, *et al*. 2001. Genet Med **3**:149-154
Technical Standards and Guidelines for Cystic Fibrosis	Supplement to the ACMG Standards and Guidelines for Clinical Genetics Laboratories (in preparation)

External monitoring

All clinical laboratories performing genetic testing must comply with general regulations under the Clinical Laboratory Improvement Amendments (CLIA) and a CLIA certification should be considered the minimum acceptable level of external monitoring. One shortcoming of having only a CLIA certification is that CLIA inspectors often have less experience in evaluating genetic testing laboratories than other certifying organizations. CLIA is in the process of upgrading its regulations regarding genetic testing. The Task Force on Genetic Testing concluded that the current CLIA requirements are insufficient to ensure quality of molecular genetic testing. Laboratories certified by the College of American Pathologists or by New York State Health Department will have undergone a more rigorous external monitoring that requires specific procedures and documentation.

The need for positive cystic fibrosis assay controls

The issue of positive controls deserves particular attention. Positive controls for the standard 25 cystic fibrosis mutations must be utilized to validate the assay and for each lot of reagents. These controls (or a subset of positive controls) are recommended to be routinely included in each assay run. However, obtaining these positive controls can be difficult. Not all of the 25 recommended cystic fibrosis mutations (Grody *et al*., 2001) are readily available through the American Type Culture Collection (ATCC, Rockville, MD www.atcc.org) or the Coriell Institute for Medical Research (Camden, NJ http:://arginine.umdnj.edu) repositories. Thus,

laboratories are left to find another source for some of the positive controls. The following are four possible sources for obtaining positive controls:

- Asking colleagues who possess validated DNA aliquots of one or more of the cystic fibrosis mutations to share a portion of their material ("clone by phone"). For such sharing to take place, consent issues may need to be considered
- Synthesizing positive controls based on sequence alterations, using well designed primers, and then validating these synthetic controls using sequence analysis (sequence analysis is the "gold standard").
- Encouraging commercial reagents manufacturers to produce sets of validated positive controls for all of the 25 recommended cystic fibrosis alleles and provide them to laboratories using their reagents.
- Encouraging a foundation (e.g., the Cystic Fibrosis Foundation), a professional organization (e.g., the College of American Pathologists), or a governmental organization (e.g., the Centers for Disease Control and Prevention) to support an initiative aimed at developing and distributing positive controls to all testing laboratories.

The first two sources listed above are currently the only available options for laboratories involved with cystic fibrosis testing, and both have important drawbacks and limitations. There is a limit, for example, to how much DNA an investigator should be asked to provide. Even when benevolent investigators are found, stumbling blocks may be encountered in fulfilling the requirements of institutional materials transfer agreements or in obtaining Institutional Review Board (IRB) approvals. The second approach listed is also limited, as it requires a greater development effort than many laboratories could manage. The third approach has appeal, since manufacturers (including those supplying Analyte Specific Reagents) must validate their product prior to introduction and must, therefore, already have access to all necessary positive controls. Lastly, organizations such as those listed could embark on a coordinated initiative aimed at developing and distributing cystic fibrosis positive controls to testing laboratories.

References

Grody WW, Cutting GR, Klinger KW, Richards CS, Watson MS, Desnick RJ. 2001. Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet Med* **3**:149-154.

Holtzman NA, Watson MS. 1997. Promoting Safe and Effective Genetic Testing in the United States. Final report of the Task Force on Genetic Testing. http://www.nhgri.nih.gov /ELSI/TFGT_final/, pp1-72.

Neumaier M, Braun A, Wagener N. 1998. Fundamentals of quality assessment of molecular amplification methods in clinical diagnosis. International Federation of Clinical Chemistry Scientific Division Committee on Molecular Biology Techniques. *Clin Chem* **44**:12-26.

ANALYTIC VALIDITY

Question 12: Have repeated measurements been made on specimens?

Summary

- Having information about repeated measurements on the same specimen is important for determining the type and rate of errors in detecting cystic fibrosis mutations
- External proficiency testing programs are the only available source of data for repeated measurements on the same specimen by multiple laboratories
- All clinical laboratories test control samples repeatedly, but results are not usually reported

Measurements made on the same specimen in different laboratories

Multiple laboratories have made repeated measurements on the same specimen, utilizing a variety of technologies. A collaborative external proficiency testing program, jointly administered by the American College of Medical Genetics and the College of American Pathologists (ACMG/CAP) provides up to six cystic fibrosis DNA challenges each year, along with a summary report of the results. An earlier section in Analytic Validity (Questions 12 and 13) provides more details about the results of this program. In the first distribution of 2000, 32 of 36 participating laboratories (89 percent) correctly reported the results for all three challenges. Two of the challenges were wild-type and one was heterozygous for a cystic fibrosis mutation; all 36 laboratories correctly classified the two wild-type specimens. In 1999, 42 of 43 participating laboratories (98 percent) correctly classified two specimens heterozygous for a common cystic fibrosis mutation, and two specimens (one heterozygous and one homozygous) for another mutation. In 1998, 51 laboratories reported their analyses of three cystic fibrosis challenges, one homozygous and two compound heterozygous specimens. The correct response rates were 88 percent, 92 percent, and 100 percent, respectively. Although there may be some explanations for the poor between-laboratory performance on some challenges (difficulties with distinguishing delI507 and delF508 have been clearly documented), these data indicate that there is also unexplained variability between laboratories when measuring the same specimen. In summary, the between-laboratory replication of a single specimen is between 90 percent and 98 percent, when detectable mutations are present.

Measurements made repeatedly on the same sample within a laboratory

It is common practice for repeated measurements to be made on the same specimen (a control specimen) within a laboratory. For each assay, a positive control is usually included for testing. This internal documentation will remain within the laboratory but will be available for on-site inspections by certifying agencies. Thus, one avenue for collection of these data would again be to use laboratory survey instruments. In one laboratory offering prenatal screening for cystic fibrosis (Knight GJ, Foundation for Blood Research, ME, personal communication), two multi-mutation controls (three in one control; two in the other) were run on 60 consecutive clinical assays in 2000 and early 2001. Overall, two failures occurred (both were the three-mutation control) for a failure rate of 2.0 percent (6/300). In another laboratory offering testing for cystic fibrosis (Roa B, Baylor University Medical Center, personal communication), the delF508 control did not fail in 52 assays in 2000. Nearly all laboratories will have these data available, even though they may not be routinely collated and analyzed.

ANALYTIC VALIDITY

Question 13: What is the within- and between-laboratory precision?

This question is not applicable to prenatal screening for cystic fibrosis, since such testing is qualitative. This question is only relevant to quantitative measurements such as repeat sizing.

ANALYTIC VALIDITY

Question 14: If appropriate, how is confirmatory testing performed to resolve false positive results in a timely manner?

Summary
- Confirmatory testing is additional testing to confirm the finding of a mutation(s)
- Such testing should be considered when a carrier, carrier couple, or affected fetus is identified
- It is likely to be useful in selected circumstances, because of occasional false positive test results
- There is little information about how often confirmatory testing corrects an error
- The type of confirmatory testing depends on the clinical circumstances, sample type and testing methodology
- Supplementary testing might occasionally be necessary

Definition

Confirmatory testing is performed to ensure that the initially positive test result is correct. For example, rerunning a specimen that was positive for G551D in order to ensure that it was correct is considered confirmatory testing. Reflexive testing is different from confirmatory testing in that other mutations or polymorphisms are being analyzed to aid in the interpretation of positive results. For example testing an unexpected homozygous delf508 individual for the presence of the benign polymorphism delF508C would be considered reflexive testing. This is also the case in testing for the 5T/7T/9T polymorphism after identifying the R117H mutation.

In prenatal screening for cystic fibrosis, confirmatory testing of some type should be considered in the following three circumstances:
- A cystic fibrosis mutation is identified in an individual
- A cystic fibrosis mutation is identified in both members of a couple
- A fetus with two cystic fibrosis mutations is identified

Four distinct types of confirmatory testing could be utilized, depending on the testing protocols in place and the circumstances in which the positive test result is obtained.
- Repeating the same test protocol on another aliquot of the same specimen
- Repeating the same test protocol on a different (or further processed)* specimen
- Performing a different test protocol on another aliquot of the same specimen
- Performing a different test protocol on a different (or further processed)* specimen

* further processing would include, for example, culturing fetal cells obtained via chorion villus sampling

Confirmatory testing in a prenatal cystic fibrosis screening program

In the following paragraphs, the four types of confirmatory testing are examined in each of the circumstances in which a mutation is detected. When a mutation is identified in a screened individual (about 1 in 29 non-Hispanic Caucasians), laboratories may repeat their testing

protocol on another aliquot of the same sample in order to rule out sample mishandling or laboratory sample mix-up. Only in rare circumstances will any other type of confirmatory testing be performed. Rare findings, such as a previously unknown, asymptomatic homozygote might warrant confirmation with another sample, possibly with a different methodology.

In the event that a laboratory identifies a mutation in both members of a screened couple (occurring about 1 in 900 non-Hispanic Caucasian couples), further confirmatory testing might also be warranted in certain circumstances. For example, if the initial specimens were mouthwash or buccal scraping, laboratories might consider collecting and analyzing a blood specimen to demonstrate chain of custody, prior to the option of invasive testing. Some laboratories may want to verify the results using a different testing protocol, or send an aliquot of the sample to a reference laboratory for confirmation. It is important to ensure accurate results, since diagnostic testing will be an option for this couple not only for this pregnancy, but also for all subsequent pregnancies.

Some laboratories perform DNA analysis for cystic fibrosis on amniotic fluid cells prior to culture, but require confirmation on cultured cells from the same fluid. When a positive test result is identified using cells obtained from the fetus (amniotic fluid, chorion villus sampling, or umbilical blood), testing may be repeated in a second laboratory to ensure accurate results, but most experienced diagnostic laboratories feel confident in their results. There are no other tests available for confirmation in the prenatal setting; the molecular test is the definitive test. It is especially critical, in this diagnostic setting, to ensure accurate results.

Importance of confirmatory testing

The analytic specificity is currently estimated to be 97.9 percent (Question 10). It is important, therefore, to determine how often 'false positive' results will be identified upon confirmatory testing. If the error is due to clerical or laboratory sample mix-up, simple retesting of an additional aliquot may be sufficient to identify and correct the error. Given that proficiency testing in Europe found 90 percent of the errors to be of this type (Dequeker and Cassiman, 2000), confirmatory testing can be expected to eliminate many of the false positive results. This issue is dealt with in more detail under Clinical Performance (Questions 19 and 20).

Other additional testing in a prenatal cystic fibrosis screening program

A supplementary test (reflexive testing for the 5T variant) has been recommended when the R117H mutation is identified (Grody *et al*., 2001). The finding of the R117H allele on a 5T background (in cis) is interpreted as a classic cystic fibrosis mutation, whereas R117H on a different background may be associated with congenital bilateral absence of the vas deferens (CBAVD) or other non-cystic fibrosis phenotypes.

The need for a second type of supplementary test arises when a delF508 allele is identified by certain types of DNA methodologies. A few analytic methods may not always distinguish between delF508 (wild type) and several benign polymorphisms (e.g., delF508C). Assaying the sample using a second test method (such as gel analysis or a method known not to cross react with these benign polymorphisms) will allow the genotype to be correctly identified and interpreted. Such testing could also be useful when an unexpected homozygous (or compound heterozygous) individual is identified.

References

Grody WW, Cutting GR, Klinger KW, Richards CS, Watson MS, Desnick RJ. 2001. Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet Med* **3**:149-154.

Cuppens H, Cassiman JJ. 1995. A quality control study of CFTR mutation screening in 40 different European laboratories. The European Concerted Action on Cystic Fibrosis. *Eur J Hum Genet* **3**:235-245.

Dequeker E, Cassiman J. 2000. Genetic Proficiency testing in diagnostic laboratories – quality control is the message. *Am J Hum Genet* **67**:A274.

ANALYTIC VALIDITY

Question 15: What range of patient specimens has been tested?

Summary
- Both whole blood and buccal lysates are acceptable for screening
- Blood samples are more expensive and require collection at a medical facility, but are associated with more generous amounts of high quality DNA.
- Buccal lysates are less expensive and can be collected at home, but are associated with smaller amounts of lower quality DNA
- Several sources of fetal DNA can be used for diagnostic testing

Cystic fibrosis mutation analysis has been successfully performed in a variety of specimens using available methodologies.

Screening can be performed on:
- whole blood (purified DNA and lysates),
- buccal lysates (cheekbrush, swab and mouthwash), or
- dried blood spots.

Diagnostic testing of the fetus can be performed on:
- direct and cultured amniotic fluid cells,
- chorionic villi collected via chorionic villus sampling (CVS),
- cells obtained via percutaneous umbilical blood sampling (PUBS)

Pre-implantation diagnostic testing can be carried out on a single cell.

Molecular confirmation of diagnosis can be performed on products of conception and tissue samples.

Blood samples are the most reliable method of collecting large amounts of high quality DNA, but a trained phlebotomist is needed, thereby increasing costs and requiring that specimens be collected at a medical facility. Buccal cells obtained by scraping, brushing or mouthwash yield adequate amounts of DNA for screening purposes (Doherty *et al*., 1996; Loader *et al*., 1996; Witt *et al*., 1996; Grody *et al*., 1997). This technique can be used to collect samples at the physician's office or at home. Buccal samples have the disadvantage of less DNA, higher failure rates, and less documentation of chain of custody. Buccal lysates can be frozen and stored for years and still be tested successfully (Bradley *et al.,* 1998). Dried blood spots can also be used for PCR-based testing, and experience has been gained in newborn pilot screening trials (Summary of newborn trials contained in Question 21). However, they have not routinely been used in prenatal cystic fibrosis pilot screening programs. In an informal survey of several commercial laboratories offering cystic fibrosis testing, none accepted blood spots (Gasparini *et al*., 1999 (S Richards, personal communication).

For diagnostic purposes, it is most difficult to analyze fetal cells prior to culturing due to limited numbers of cells and the high potential for maternal cell contamination. When fetal specimens are tested, it is the laboratory's responsibility to assess maternal cell contamination.

References

Bradley LA, Johnson DD, Palomaki GE, Haddow JE, Robertson NH, Ferrie RM. 1998. *J Med Screen* **5**:34-36.

Doherty RA, Palomaki GE, Kloza EM, Erickson JL, Haddow JE. 1996. Couple-based prenatal screening for cystic fibrosis in primary care settings. *Prenat Diagn* **16**:307-404.

Gasparini P, Arbustini E, Restagno, Zelante I, Stanziale P, Gatta L, Sbaiz L, Sedita AM, Banchieri N, Sapone L, Fiorucci GC, Brinson E, Shulse E, Rappaport E, Fortina P. 1999. Analysis of 31 CFTR mutations by polymerase chain reaction/oligonucleotide ligation assay in a pilot screening of 4476 newborns for cystic fibrosis. *J Med Screen* **6**:67-69.

Grody WW, Dunkel-Schetter C, Tatsugawa ZH, Fox MA, Fang CY, Cantor RM, Novak JM, Bass HN, Crandall BF. 1997. PCR-based screening for cystic fibrosis carrier mutations in an ethnically diverse pregnant population. *Am J Hum Genet* **60**:935-947.

Loader S, Caldwell P, Kozyra A, Levenkron JC, Boehm CD, Kazazian HH Jr, Rowley PT. 1996. Cystic fibrosis carrier population screening in the primary care setting. *Am J Hum Genet* **59**:234-247.

Witt DR, Schaefer C, Hallam P, Wi S, Blumberg B, Fishbach A, Holtzman J, Kornfeld S, Lee R, Nemzer L, Palmer R. 1996. Cystic fibrosis heterozygote screening in 5,161 pregnant women. *Am J Hum Genet* **58**:823-835.

ANALYTIC VALIDITY

Question 16: How often does the test fail to give a useable result?

Summary

- Laboratory testing for cystic fibrosis mutations can be divided into pre-analytic, analytic and post-analytic phases
- In the pre-analytic phase, generally agreed upon criteria are in use to determine the appropriateness of testing. If these are not met, the test can be canceled
- In the analytic phase, samples fail for multiple reasons, and these failures are routinely documented in clinical laboratories but are not generally available for outside review
- When analytic failures do occur, repeating the analysis will often yield useable results
- Types of failures and their associated rates are rarely reported as part of pilot trials or method comparisons

Test 'failures' in the pre-analytic phase of testing

In the pre-analytic phase, it may be determined that the sample is not suitable for testing because specific clinical criteria are not met, or because the sample is considered inadequate. While programs often monitor pre-analytic test cancellation rates as part of an overall quality assurance plan, these events are usually not considered a laboratory or methodologic 'failure'. Table 2-9 lists criteria commonly used for deciding whether to reject a sample in the pre-analytic phase.

Table 2-9. Common Pre-analytic Criteria for Rejecting a Sample Submitted for Prenatal Cystic Fibrosis Screening

Rejection Criteria Based on Clinical Information

Gestational age too advanced
(e.g., received after 20 or 21 weeks' gestation)
Testing already performed on a previous sample for this couple
Couple has a family history of cystic fibrosis
(more extensive DNA testing may be indicated)

Rejection Criteria Based on Submitted Sample

Inadequate specimen quality
(e.g., hemolyzed blood, dried buccal sample or obvious contamination)
Inappropriate sample
(e.g., whole blood with no anticoagulant or wrong anticoagulant)
Inadequate specimen labeling
Inappropriate handling prior to laboratory receipt
(e.g., sample too long in transit or exposed to extreme temperature)

Test failures during the analytic phase of testing

Failures of individual samples or assays occur when preset quality control standards are not met and test results are not reportable. Failures can arise for a number of reasons such as improperly

processed samples, problems with component reagents, or equipment malfunction. Many assay failures within the clinical molecular genetic laboratory are due to operator error. Automation and programs to properly train laboratory personnel can avoid most of these problems. Only a few medical technology programs, however, currently provide adequate molecular components in their programs. Documentation of failures and subsequent corrective action is required by regulatory agencies such as CLIA and the College of American Pathologists. Unfortunately, failure rates and other information on assay robustness are often not published as part of pilot trials or method evaluations. Available data suggest, however, that repeating the analysis of an individual sample or assay run can often yield a satisfactory result.

A irretrievable assay failure occurs when an apparently suitable specimen is submitted and approved for testing, but the assay yields a result that is clinically uninterpretable. Failures of this type are most often related to the quality of the original sample. Procedural problems during specimen processing and DNA extraction can also be responsible. Success rates for obtaining clinically interpretable results are close to 100 percent for blood samples. Buccal samples have a somewhat lower success rate (98 percent to over 99 percent) as a result of poor sampling (inadequate number of cells), sample contamination, desiccation (exposure to extreme heat), or inadequate sensitivity of the testing methodology to account for the lower concentration and quality of the sample.

Post-analytic failures, such as incorrect or inadequately interpreted results are considered separately from analytic test failures, as part of a review of overall quality assurance in the Clinical Utility section (Question 32).

Gap in Knowledge: Overall, and method-specific failure rates
Clinical laboratories are required to document test failures as described above. For this reason, this type of information should be readily available from laboratories participating in external proficiency testing administered by the ACMG/CAP. This could be accomplished though the use of a supplemental question attached to a routine distribution or, alternatively, the data could be collected via an externally funded, independent project.

ANALYTIC VALIDITY

Question 17: How similar are results obtained in multiple laboratories using the same, or different technology?

Summary

- Data derived from external proficiency testing can be used to judge the consistency of results from different cystic fibrosis screening laboratories
- Stratification of results by methodology does not currently yield reliable information because of the small number of laboratories participating in proficiency testing and the large number of methodologies,
- Overall, the results from multiple laboratories appear to be similar, regardless of the methodology used, if the panel of mutations employed by individual laboratories is taken into account.

Comparing results from different laboratories using the same or similar methodologies
The only potential source of data for evaluating differences in cystic fibrosis test results from multiple laboratories using the same (or a similar) method would be derived from external proficiency testing. However, the relatively small number of participants and the relatively large number of methods (Table 2-10, Appendix B) preclude obtaining meaningful method-specific analyses. Even if available, such comparisons might be complicated because laboratories in the same methodological category may be using different commercial or in-house reagent components and protocols. For example, although three laboratories might be grouped under the ARMS™ methodology, one might use a prepared kit, a second might use commercially prepared ASR's (analyte specific reagents), and the third might use in-house reagents. Each may also be targeting a different set of mutations. All of these factors would make the comparison nearly equivalent to comparing different methodologies. To help in comparing methodologies, the ACMG/CAP MGL Survey Reports might consider stratifying results into broad methodological categories.

Comparing results from different laboratories regardless of the methodology
As part of the 2000 ACMG/CAP Molecular Genetics Laboratory external proficiency testing survey, laboratories were queried about their methodology for performing cystic fibrosis mutation analysis (Table 2-10, Appendix B). Overall, the reported methodologies were used to detect between 1 and 70 mutations (median 12 to 14 mutations). To date, method-specific data on error rates are not available from these surveys. However, during the six years of operation (1996 through 2000) there was a high level of agreement between laboratories for detecting mutations that were targeted by their specific method.

The European Concerted Action on Cystic Fibrosis reported results from a much larger number of laboratories using surveys in 1996 through 1998 (136, 145 and 159 laboratories, respectively). Again, few method-specific data are available. However, all commercial kits were found to have problems identifying G551D and R553X.

The impact of a laboratory's mutation panel on proficiency testing results

There were instances when the responses varied greatly, because of variability of mutations being tested for by laboratories. For example, in 2000 MGL-17, DNA from a compound heterozygote (621+1G>E / G85E) was distributed.

- Twenty of the 41 participating laboratories (49 percent) obtained the correct genotype.
- Sixteen other laboratories (39 percent) did not test for G85E and thus identified the sample as coming from a heterozygote (621+1G>E/wild).
- Four other laboratories (10%) did not test for either mutation and reported a normal genotype (wild/wild).
- One laboratory (2 percent) did not test for 621+1G>E and thus reported a heterozygote (G85E/wild).

Although this genotype is rare, it demonstrates the wide range of laboratory responses that can occur when a mix of methodologies and mutation panel sizes occurs in practice. A more complete discussion of mutation panel size, composition and performance is contained in Clinical Validity.

Gap in Knowledge: Comparison of Methods for Cystic Fibrosis Mutation Detection

In order to compare analytic validity for various testing methodologies, proficiency testing data need to be stratified by methodological category. It would also be useful to identify subsets using the same commercially available reagents (e.g., in-house versus ASR reagents). Alternatively, a previously described method for validation (Question 9 and 10 – Optimal Sources of Data) could be employed that would provide not only analytic performance for a methodology, but also comparative data between methodologies.

Appendix B. Analytic methodologies used for cystic fibrosis mutation analysis

Introduction

Table 2-10 lists categories of methodologies that are used to detect cystic fibrosis mutations by laboratories participating in proficiency testing programs in the United States (ACMG/CAP MGL Survey) and Europe, along with the proportions using each method. Because many laboratories utilize "home brew" assays, these categories are not homogeneous. Some methodologies are relatively labor intensive and can only detect a few mutations (e.g., heteroduplex analysis), making them more suitable for research or diagnostic laboratories. When large numbers of specimens must be tested with short turn-around times (e.g., prenatal screening), other methodologies are needed. The European report documents a clear increase in use of commercially prepared materials/kits, 28 percent of laboratories in 1996 versus 50 percent in 1998 (Dequeker and Cassiman, 2000).

Table 2-10. Testing Methods Utilized by 36 US Laboratories and 151 European Laboratories According to External Surveys

Testing Method	USA (%)[a]	Europe (%)[b]
Allele Specific Oligonucleotide (ASO)	39	24
Electrophoresis for RFLP and size analysis	39	2
Allele-specific PCR/ARMS	19	15
Oligonucleotide ligation analysis (OLA)	14	18
All Mutation Scanning Methods	11	41
(Heteroduplex analysis)		(36)[c]
(SSCP)		(1)
(DGGE)		(2)
(Other)		(2)
Other methods	28	0

[a] ACMG/CAP 2000 MGL Totals more than 100 percent (some laboratories use more than one methodology)
[b] European Concerted Action for Cystic Fibrosis 1998
[c] Mutation scanning methods available only as a total percent for USA. Numbers in parentheses delineate individual methods for Europe

Cystic fibrosis testing methodologies for prenatal screening ought to include the following characteristics:

- a reasonable number of mutations
- a low to moderate level of technical expertise
- a short turn-around time (one or two days)
- a high throughput (ideally, on an automated platform)
- a relatively low cost

These requirements might appear ambitious, but the evolution of other tests now used for screening in the clinical laboratory shows these goals are achievable. For example,

immunoassays that are now routinely performed were originally developed in the 1960's by investigators with in-depth knowledge of immunochemistry and radiation detection methods. Over the ensuing years, these assays were revised and streamlined by manufacturers to meet the needs of clinical laboratories, including the development of automated immunoassay systems that minimize the chance for error. For FDA approved kits, the responsibility for ensuring reagent quality and instrument performance now rests primarily with the manufacturer. The laboratory's responsibility is to monitor the quality control measures set by the manufacturer to verify that assay performance meets specifications. A further development is a computer link to the instrument that automatically transfers test results to a patient record system for reporting. Automation is more expensive than manual assays in terms of reagents and instrument rental or purchase, but the overall cost per test can be the same or lower because of the reduced labor costs. This same development is beginning to occur for prenatal cystic fibrosis screening. Commercially prepared reagents have emerged using three major methodologies. The attributes of these reagents are summarized in Table 2-11.

Table 2-11. Characteristics of Commercial Reagents to Detect Cystic Fibrosis Mutations

	Commercial Cystic Fibrosis Mutation Detection Systems			
Characteristic	**Elucigene**	**INNO-LiPA**	**Linear Array**	**PCR OLA**
1. Method Type	ARMs	ASO	ASO	OLA
2. Company	Orchid	Innogenetics	Roche Molecular	Perkin Elmer
3. Mutations	29	33	31	31
4. Robustness	High	High	Not available	High
5. Special equipment	No	No	No	Yes
6. Total time (days)	1 to 1.5	0.5 to 1	2	0.5 to 1
7. Cost per patient	$30-40	$30-40	Not available	$55
8. Advantages	Low complexity Published data	Can be automated	Can be automated	Can be semi-automated
9. Disadvantages	Cannot now be automated		Automation essential for high throughput	Many steps
For more information	www. orchid.com	www. innogenetics.com	http://biochem. roche.com	www.applied biosystems.com

Notes pertaining to Table 2-11:

1. Method type: Methods displayed are those that are most commonly used and that are suitable for large-scale cystic fibrosis screening. These include the Amplification Refractory Mutation System (ARMS™), Oligoligation assay (OLA), and allele specific oligonucleotide assays (ASO). Both of the ASO assays use reverse dot blot strip technology. For more information about these methodologies, including a description and set of references, see www.ich.ucl.ac.uk/cmgs/.

2. Company: None of these commercial reagents have been approved by the FDA for clinical use. However, reagents may qualify under the FDA's Analyte Specific Reagent (ASR) rule which indicate that the assay building blocks are made under good manufacturing practices.

3. Mutations: This is the number of cystic fibrosis mutations that can be detected by the testing protocol. Manufacturers are modifying existing reagents to conform with the panel of 25 mutations. The EluciGene ARMS™ test cannot reliably identify individuals who carry two copies of a mutation other than delF508 (about 2-3 percent of affected individuals). This is not a critical requirement in population screening, where carriers are the initial target. Other methods can identify both carriers (heterozygotes) and homozygotes.

4. Robustness: Robustness describes how consistently and reliably a set of reagents performs when used by different laboratories, under varying conditions, and on different sample types (e.g., blood, buccal smears).

5. Special equipment: Some manufacturers require that specialized equipment be used to perform their assays. Although initially more costly, the equipment may allow more samples to be tested.

6. Total time: Estimated time to complete assay, including sample processing and reporting. Some methods only require one day, but laboratories may choose to extend the process to a second day for more convenient scheduling.

7. Cost per patient: Costs for the reagents and licenses to perform cystic fibrosis testing are extremely variable. Some laboratories perform 'in-house' assays with relatively low reagent costs. In such cases, the cost of technical time for reagent preparation and QC/QA must also be considered. Costs of analyte specific reagents (ASR) can be relatively high, compared to traditional biochemical assays. However, the savings in technical staff time for preparation and QC/QA can offset reagent costs. For screening, the relevant figure is the cost per patient tested, rather than the cost per mutation tested.

8. Advantages: Reagents for prenatal cystic fibrosis screening should have high throughput with relatively low labor costs. Assays that can be efficiently automated can be cost effective. Peer-reviewed analytic validity data are helpful for validation.

Newer testing technology platforms with high potential for cystic fibrosis testing include various hybridization strategies (Roche and Luminex), arrayed primer extension (Orchid), mass spectrometry (Sequenom), sequence analysis (Pyrosequencing), and random coated array detection (Bioarray Solutions). However, there are no existing data that accurately compare these technologies with currently utilized methodologies or with each other.

References

Dequeker E, Cassiman JJ. 2000. Genetic testing and quality control in diagnostic laboratories. *Nat Genet* **25**:259-260.