

United States Court of Appeals
FOR THE DISTRICT OF COLUMBIA CIRCUIT

Argued March 29, 1996 Decided May 21, 1996

No. 95-5384

CONSUMER FEDERATION OF AMERICA AND
PUBLIC CITIZEN,
APPELLEES

v.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES,
APPELLANT

Appeal from the United States District Court
for the District of Columbia
(No. 93cv00097)

John S. Koppel, Attorney, United States Department of Justice, argued the cause for appellant, with whom *Frank W. Hunger*, Assistant Attorney General, *Michael J. Singer*, Assistant Director, and *Eric H. Holder, Jr.*, United States Attorney, were on the briefs.

Allison M. Zieve argued the cause for appellees, with whom *Brian Wolfman* was on the brief.

Before: WALD, SILBERMAN, and SENTELLE, *Circuit Judges*.

Opinion for the court filed by *Circuit Judge* WALD.

WALD, *Circuit Judge*: In 1992, the United States Department of Health and Human Services ("HHS" or "agency") issued regulations implementing the 1988 Clinical Laboratory Improvement Amendments ("CLIA" or "Act"), which expands and strengthens federal regulation of laboratories that perform clinical tests on specimens from the human body. The specific regulations at issue here pertain to two key components of CLIA: first, a regime for classifying laboratory tests and developing uniform national personnel standards applicable to each of these test categories, and second, a proficiency testing program designed to ensure adequate performance by cytologists¹ working in clinical labs. Consumer Federation and Public Citizen, two national organizations

¹"Cytology is the examination of cells [under a microscope] to identify abnormalities which may indicate disease." H.R. REP. NO. 899, 100th Cong., 2d Sess. 16 (1988) [hereinafter H.R. REP. NO. 899], *reprinted in* 1988 U.S.C.C.A.N. 3828, 3837.

concerned with consumer and health advocacy, challenged the HHS regulations as contrary to the dictates of CLIA. Ruling on cross-motions for summary judgment, the district court invalidated both elements of the CLIA regulatory scheme and ordered HHS to engage in expedited rulemaking on remand.

HHS now appeals the district court's ruling, seeking reinstatement of its regulations, or in the alternative, an order vacating the district court's instructions to expedite new rulemaking. Because we agree with the agency that its regulations regarding personnel qualifications represent a permissible construction of the relevant CLIA provisions, we reverse the district court's invalidation of the personnel regulations. As to the regulations covering cytology proficiency testing, we find that while the agency's belated explanation for its actions in a declaration submitted by an agency official to the district court might have passed muster if timely proffered, that declaration constitutes *post hoc* reasoning not included in the agency record. Since review of an agency action must be based on the administrative record alone, we do not consider this supplementary information and accordingly remand the proficiency testing regulations to the agency to develop an adequate explanation on remand, if it can do so.

I. BACKGROUND

In response to growing public concern about the quality of clinical laboratory testing, Congress passed CLIA in 1988 to ensure competence among laboratories testing human specimens for disease diagnosis, prevention, monitoring, and treatment. Prior to CLIA, the federal government regulated only those laboratories which received reimbursement from the federal Medicare program or handled test samples shipped in interstate commerce. H.R. REP. NO. 899 at 11, *reprinted in* U.S.S.C.A.N. at 3831; S. REP. NO. 561, 100th Cong., 2d Sess. 3-4 (1988) [hereinafter S. REP. NO. 561]. All other labs performing clinical tests were free from federal regulation and oversight. Following extensive hearings and investigations, Congress concluded that some of these unregulated facilities "pose[d] a serious threat to the public health." H.R. REP. NO. 899 at 14, *reprinted in* U.S.S.C.A.N. at 3835. Because they processed tests without sufficient regard for accuracy or reliability, these labs reported incorrect test results at an unacceptably high rate. *Id.*; S. REP. NO. 561

at 4-5; 27. In particular, significant attention was focused on laboratories reviewing Pap smears, which are used to screen women for indications of cervical cancer. Witnesses presented evidence that many labs were employing inadequately trained cytologists and requiring them to process tests at rates two to three times higher than the recommended maximum workload. S. REP. NO. 561 at 5. As a consequence, the labs reported large numbers of false negative results, contributing to unnecessary suffering and even death in women who did not receive prompt treatment for cervical cancer because the labs failed to identify their Pap smears as abnormal. H.R. REP. NO. 899 at 16-17, *reprinted in* U.S.S.C.A.N. at 3837-38; S. REP. NO. 561 at 27.

Congress identified a number of reasons for these deficiencies, including the absence of uniform standards for personnel performing clinical tests, rates for screening cytology specimens which were so rapid that the risk of improper diagnosis was extremely high, ineffective proficiency testing (or none at all), and the lack of proper quality control measures. H.R. REP. NO. 899 at 12-17, *reprinted in* U.S.S.C.A.N. at 3833-38; S. REP. NO. 561 at 4-5. To remedy these critical shortcomings, Congress expanded HHS's oversight responsibilities to include all laboratories performing clinical tests on human specimens and revamped the existing regulatory scheme. Two of these reforms are at issue here: a system of personnel qualifications, and proficiency testing requirements for cytologists.

CLIA directs HHS to establish personnel qualifications for all individuals performing clinical tests, and further instructs that the qualifications "shall, as appropriate, be different on the basis of ... the risks and consequences of erroneous results associated with such examinations and procedures." 42 U.S.C. § 263a(f)(1)(C) (1994). As explained in greater detail below, HHS divided all clinical tests into three categories, and assigned increasingly stringent personnel qualifications to each category. Although it had initially classified the tests according to, *inter alia*, the consequences of an erroneous test result, it dropped this factor from the classification scheme in its final rule. Instead, the agency categorized tests on the basis of several other factors, the most prominent of which was the complexity of the testing procedure, and imposed increasingly demanding personnel standards for each successively complicated category.

CLIA also requires periodic proficiency testing of cytologists, to be conducted "to the extent practicable, under normal working conditions." *Id.* § 263a(f)(4)(B)(iv). Despite this language, the testing protocol selected by HHS used a work rate significantly lower than the maximum permissible daily work rate. While a cytologist could screen up to 100 slides per day at work (a work rate of 12.5 slides per hour), the regulation indicated that testing should take place at a rate of 5 slides per hour, and also stipulated that a proficiency test include a much higher proportion of abnormal slides than would occur in the average work day (30% to 60% abnormal slides, as compared with a workday average of 5% abnormal slides). 42 C.F.R. §§ 493.855, 493.945 (1995).

Consumer Federation of America ("Consumer Federation") and Public Citizen² challenged both parts of the CLIA regulations as arbitrary, capricious, and in violation of law. On cross-motions for summary judgment, the district court ruled for the plaintiffs, finding that neither regulation comported with the relevant statutory requirements. Although it permitted the challenged regulations to remain in place pending issuance of a new rule, the district court ordered HHS to promulgate new proposed regulations within 90 days, and a final rule "within a reasonable time thereafter." *Consumer Fed'n of America v. Dep't of Health and Human Services*, No. 93-97, Order at 2 (D.D.C. Aug. 29, 1995). HHS then requested a partial stay of the portion of the district court's order requiring the agency to expedite issuance of a new proposed rule regarding personnel qualifications. Although the district court denied the motion, it was granted by this court on December 1, 1995. HHS now appeals the district court's ruling as to its regulations on both personnel qualifications and cytology proficiency testing, arguing that the challenged regulations are in accordance with the relevant provisions of CLIA. In the alternative, the agency seeks reversal of the district court's order to engage in expedited rulemaking.

II. ANALYSIS

We now turn to each set of regulations successfully challenged by Consumer Federation in the district court. We review *de novo* the district court's order granting Consumer Federation's motion for summary judgment and denying the government's cross-motion for summary judgment.

²For simplicity, these two organizations are hereinafter referred to as "Consumer Federation."

Center for Auto Safety v. Federal Highway Admin., 956 F.2d 309, 312 (D.C. Cir. 1992). Summary judgment is appropriate only when the record reveals no genuine issue of material fact. *See* FED. R. CIV. P. 56(c); *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 248 (1986). Because the contested regulations involve two different provisions of CLIA, we treat them separately, considering the relevant statutory language, regulatory background, and procedural history of each one.

A. *Personnel Qualifications*

The section of CLIA directing HHS to develop national personnel standards requires that medical labs

use only personnel meeting such qualifications as the Secretary may establish ... which qualifications shall, as appropriate, be different on the basis of the type of examinations and procedures being performed by the laboratory and the risks and consequences of erroneous results associated with such examinations and procedures.

42 U.S.C. § 263a(f)(1)(C). In its May, 1990 Notice of Proposed Rulemaking ("NPRM") implementing this provision, HHS proposed a three-tiered scheme of personnel qualifications. Each type of lab test would be placed into one of three categories—"waived" tests (those so simple to perform that the likelihood of an erroneous test result is extremely small, as well as those which pose no reasonable risk of harm to the patient if performed incorrectly), Level I tests, or Level II tests—depending upon the consequences to the patient if the test was performed incorrectly, the complexity of the testing methodology, the degree to which the test requires independent judgment and interpretation, the degree to which interpretation of the test result requires knowledge of external variables, and the training required to perform the test. Regulations Implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA '88), 55 Fed. Reg. 20,895, 20,901-02 (1990). Under this scheme, personnel would have to satisfy increasingly rigorous qualifications to perform waived tests, Level I tests, and Level II tests, respectively. *Id.* at 20,902.

During the notice-and-comment period, HHS received approximately 14,470 comments regarding this classification model, more than 95% of which expressed opposition to the scheme. *See* Regulations Implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA), 57 Fed. Reg. 7002, 7016 (1992). In general, opponents described the model as "essentially unworkable" because it failed to realistically represent actual testing patterns or account for the multitude of testing

methodologies and instruments in use. *Id.* Several commenters focused on the criterion which measured the consequences to a patient of an erroneous test result (the "risk of harm" criterion), contending that it was "too vague and indeterminate" to be consistently applied. See *id.* at 7020.

In the final rule, the agency retained a three-tiered system, but substantially changed the way in which tests were classified by eliminating the "consequences of erroneous test result" criterion and focussing primarily on the complexity of the test procedure.³ It set forth three categories of tests: "waived," "moderate complexity," or "high complexity." As in the NPRM, "waived tests" included those which are so simple to perform that the likelihood of an erroneous test result is extremely small.⁴ These tests were exempt from regulation. *Id.* at 7019-20. The remainder of all clinical tests were assigned a score of one to three, for each of seven different criteria: (1) knowledge needed to perform the test; (2) training and experience needed to perform the test; (3) complexity of reagent and materials preparation; (4) characteristics of the steps required to perform the test; (5) availability

³The final rule made an additional change relevant to this appeal. Instead of classifying tests in broad groups (*e.g.*, all tests for performing red blood cell counts would be classified as one group), as did the NPRM, it classified each test separately (*e.g.*, individual classification of each different test used for red blood cell counts). Compiled List of Clinical Laboratory Test Systems, Assays, and Examinations Categorized by Complexity, 58 Fed. Reg. 39,860, 39,864 (1993). Thus, the total number of tests in the classification scheme increased significantly; for example, although Level I included only eleven test categories in the NPRM, its companion Level I in the final rule included thousands of tests. *Id.* at 39,879-90, 39,897-39,942.

⁴The CLIA provision governing waived tests describes them as

simple laboratory examinations and procedures which, as determined by the Secretary, have an insignificant risk of an erroneous result, included those which (A) have been approved by the Food and Drug Administration for home use, (B) employ methodologies that are so simple and accurate as to render the likelihood of erroneous results negligible, or (C) the Secretary has determined pose no reasonable risk of harm to the patient if performed incorrectly.

42 U.S.C. § 263a(d)(3). In its final rule, the agency stated that "[t]he primary criterion for placing a test on the waived list is that the test is so simple to perform ... that the likelihood of an erroneous result is extremely small." The agency also noted that "there is no test which carries with it absolutely no risk of harm if performed erroneously.... However, we do not feel that the tests on the certificate of waiver list present an insignificant risk of an erroneous result and, therefore, are exempt." 57 Fed. Reg. at 7019. While this language is sufficiently muddy that we cannot deduce whether HHS is, in fact, entirely deleting risk of harm from its criteria for inclusion on the waived list, we understand the waiver criteria to focus primarily on test complexity, and only secondarily (if at all) on the risk of harm to the patient. Regardless, the final rule expressly abandons any consideration of this factor in classifying all other tests.

of materials for calibration, quality control, and proficiency testing; (6) troubleshooting and equipment maintenance required; and (7) degree of interpretation and judgment required. See 42 C.F.R. § 493.17. In contrast to the NPRM, the agency eliminated any express consideration of the risks or consequences to the patient of an erroneous test result. 57 Fed. Reg. at 7020-21. After these values were assigned and totaled, any test with a cumulative score of 12 or less was placed in the "moderate complexity" category, and all other tests (with a score of greater than 12) in the "high complexity" category. 42 C.F.R. § 493.17. HHS then assigned specific personnel qualification requirements to labs in each category. Facilities performing only waived tests must follow accepted laboratory practices and comply with other relevant federal, state, and local requirements, but are not subject to the new CLIA standards. Labs performing moderate and high complexity tests must comply with CLIA regulations regarding proficiency testing, patient test management, quality control and assurance, and personnel standards; the standards for high complexity labs are even stiffer than those imposed on moderate complexity facilities. 57 Fed. Reg. at 7109.

Consumer Federation challenged this personnel qualification scheme before the district court. It argued that HHS's decision not to explicitly consider the risks and consequences to the patient of an erroneous test result contravened CLIA's statutory directive to establish qualifications which "shall, as appropriate, be different on the basis of ... the risks and consequences of erroneous results." Because Congress used the word "shall," Consumer Federation contended, the agency was mandated as a matter of law to expressly include the risk and consequence of error factors in its system of personnel standards. The district court agreed, reasoning that while the words "as appropriate" suggested that the agency had discretion in deciding whether to expressly incorporate the risks and consequences of error, the use of the word "shall" provided a stronger indication to the contrary that Congress meant to require explicit consideration of both factors. *Consumer Fed'n of America*, No. 93-97, Memorandum Opinion at 17-18 (D.D.C. Aug. 29, 1995) ("Mem. Op."). Accordingly, it granted summary judgment for Consumer Federation on this issue.

At the outset of our analysis, we underscore that this part of the case presents a paradigmatic *Chevron* question: do the agency's implementing regulations reflect a permissible construction of

CLIA? Although the district court did ultimately discuss whether the final rule contradicted any relevant portions of CLIA, it never expressly engaged in the two-step statutory analysis required by *Chevron*.⁵ We now proceed to do so. In reviewing an agency's construction of a statute, we first ask whether Congress has spoken unambiguously to the precise issue at hand. If it has, we give effect to Congress' intent. If not, we consider the agency's action under "Step Two" of *Chevron*, and defer to the agency's interpretation if it represents a "permissible construction" of the statute. *Chevron U.S.A. Inc. v. Natural Resources Defense Council, Inc.*, 467 U.S. 837, 842-43 (1984).

The provision at issue directs the Secretary to establish personnel qualifications which

shall take into consideration competency, training, experience, job performance, and education and which qualifications shall, as appropriate, be different on the basis of the type of examinations and procedures being performed by the laboratory and the risks and consequences of erroneous results associated with such examinations and procedures.

42 U.S.C. § 263a(f)(1)(C) (emphasis added). Consumer Federation argues that use of the term "shall" makes explicit consideration of the risks and consequences of erroneous results mandatory; the qualifying words "as appropriate" merely give the agency discretion as to *the manner* in which it considers these two factors. The government, on the other hand, contends that the statutory intent is not so clear at all. We find the government's identification of some ambiguity to be correct. Although the word "shall" is often used to impose a mandatory duty, *see Train v. City of New York*, 420 U.S. 35, 47-48 (1975); *Moon v. Dep't of Labor*, 727 F.2d 1315, 1318-19 (D.C. Cir. 1984), the inclusion of the words "as appropriate" directly following "shall" suggests that the agency is not required to include an explicit consideration of risks and consequences of error in its qualifications regime. Even without the "as appropriate," the agency still would have enjoyed discretion to decide exactly how to incorporate risks and consequences of error in its regulations, provided that it did so in some manner. If "as appropriate" is to have any effect, then, it must mean that the agency must

⁵Parsing § 263a(f)(1)(C), the district court found the language "mudd[y]" and "mutually contradictory." It then invalidated the agency's rule because it did not reflect "the *most* reasonable interpretation of that phrase, 'shall, as appropriate.'" Mem. Op. at 17-18 (emphasis added). If Congress' intent is "muddy"—in other words, ambiguous—as we agree it is, a reviewing court must defer to the agency's construction of the statute so long as it represents a *permissible* interpretation. The court need not determine that the agency's construction is the *most* reasonable one.

specifically include the risks and consequences factors in its regulations only to the extent appropriate. To conclude otherwise, as Consumer Federation advocates, would violate a basic canon of statutory construction by treating the two words as surplusage. *See Babbitt v. Sweet Home Chapter of Communities for a Great Oregon*, 115 S. Ct. 2407, 2413 (1995).⁶

Concluding that Congress has not spoken clearly to the question of whether CLIA requires HHS to explicitly consider the risks and consequences of error in formulating personnel qualifications, we proceed to the *Chevron* II analysis of whether the scheme of personnel standards established by the agency reflects a permissible construction of the statute. Here, the two factors—risks of error and consequences of error—are best considered separately, since HHS has offered different explanations for its treatment of the two criteria. As to the risks of incorrect test results, the agency contends that it did address this factor in its final rule, although not explicitly. In its 1993 rule classifying all of the clinical tests, the agency stated that the various measures of test complexity included in the classification model serve as a proxy for risk of error, because error is more likely to occur as the complexity of a test and the independent judgment required of the analyst increases. *See infra* at 8-9; 57 Fed. Reg. at 7020-21; 58 Fed. Reg. at 39,864. Consumer Federation has presented no evidence disputing this relationship between test complexity and risk of error, and we find the agency's explanation a reasonable one. Accordingly, we conclude that the agency's treatment of the risks of erroneous test results in its final rule represents a permissible construction of the relevant CLIA provision, contrary to the district court's conclusion.

The agency's admitted decision to eliminate consequences of error from classification scheme presents a knottier problem, since HHS does not claim to address this factor either directly or through use of a proxy criterion. By way of explanation, HHS indicated that it was deleting the consequences

⁶Moreover, had Congress intended to require explicit consideration of risks and consequences, it presumably would have drafted the statute differently—for example, "qualifications ... shall be different on the basis of ... the risks and consequences..." In fact, the legislative history suggests that Congress considered this option, but discarded it. The House version of CLIA used only the term "shall," while the Senate bill used the word "may," but both sides eventually compromised on the phrase, "shall, as appropriate." *Compare* H.R. 5150, 100th Cong., 2d Sess. (1988), *reprinted in* 134 CONG. REC. 23603 (Sept. 13, 1998), *with* S. 2477, 100th Cong., 2d Sess. (1988), *reprinted in* S. REP. NO. 561 at 13.

of error criterion in response to rulemaking comments about its vagueness; these comments led HHS to conclude that such a criterion was "unworkable." In its subsequent final rule, HHS elaborated upon this decision:

... [T]he consequences to the patient of an erroneous test result will vary tremendously depending on such factors as the patient's medical condition, the purpose for which a test is being conducted, and the treatment prescribed by a physician due to the test result. For example, the harm to the patient caused by an erroneous lymphocyte count will vary depending on the actual medical condition of the patient. If a serious medical condition such as leukemia goes undetected for a long period of time due to the erroneous result, then the harm to the patient may be quite serious. If however, the patient has a viral upper respiratory infection, a disease for which there is very little treatment, the consequences to the patient will be far less serious. The risk of harm will also vary depending on how a physician reacts to an erroneous test result. If an inaccurate test report leads a physician to order additional tests, then the patient will suffer no tangible harm. Incorrect test results that lead a physician to prescribe more intensive treatments, however, may have more serious consequences for the patient.

Thus, in order for the categorization process to truly reflect the risk of harm to the patient if a test is performed incorrectly, each test would have to be separately categorized based on why the test was being prescribed, the type of condition that was being tested, and the condition of the patient. Adding this layer of complexity to what was already an intricate system would have been an impossible task. Even if a classification scheme incorporating risk of harm could have been developed, the application of that scheme would have been unworkable. Under such a scheme, clinicians and laboratory directors would required [sic] to ascertain the context of each tests [sic] before determining which laboratory personnel could perform it. Introducing this type of subjectivity into the process would frustrate our goal of developing manageable regulations that would contribute to improved performance of the nations's [sic] clinical laboratories.

58 Fed. Reg. at 39,864.

The agency's rationale can be summed up as follows: although it considered inclusion of a "consequences of error" criterion in the proposed rule, upon reflection it determined that this factor can only be evaluated in hindsight and is too time-consuming to apply to each of the thousands of clinical tests. On balance we think this conclusion represents a reasonable construction of the statutory requirement that qualifications "shall, as appropriate, be different on the basis of ... [the] consequences of erroneous results." The statutory provision does require that the agency give due deliberation to the question of whether to incorporate consequences of error in its classification scheme. It does not, however, mandate inclusion of this factor if it has reasonably been found not to be appropriate. HHS has provided a coherent explanation for why this criterion is not a useful or

manageable one. These difficulties are only increased by the agency's decision to classify each clinical test separately instead of grouping tests that perform the same function together, since this methodology requires the categorization of thousands more tests than if they were categorized in groups. Legislative history further supports HHS's course of action. The committee reports accompanying CLIA do not provide much guidance on precisely how HHS shall establish personnel qualifications, but do conclude that since Congress had only limited information available to it on the connection between personnel standards and quality control, "the Secretary should be given latitude in determining both when personnel standards are needed and what those standards should be." H.R. REP. NO. 899 at 28, *reprinted in* U.S.S.C.A.N. at 3849; *see also* S. REP. NO. 561 at 24-25 (discretion granted to Secretary in establishing performance standards so that the degree of regulatory oversight may be tailored to the type of tests which a laboratory performs). We therefore find its interpretation of § 263a(f)(1)(C) a permissible one and reverse the district court's invalidation of the relevant HHS regulation.

B. *Cytology Proficiency Testing*

Turning to the second regulation challenged by Consumer Federation and invalidated by the district court—the one regarding cytology proficiency testing—CLIA requires HHS to set a limit on the rate at which cytologists may process slides, in order to prevent labs from forcing their employees to review them too rapidly. H.R. REP. NO. 899 at 31, *reprinted in* U.S.S.C.A.N. at 3852; S. REP. NO. 561 at 27. It also directs the agency to establish standards for

periodic confirmation and evaluation of the proficiency of individuals involved in screening or interpreting cytological preparations, including announced and unannounced on-site proficiency testing of such individuals, with such testing to take place, *to the extent practicable, under normal working conditions.*

42 U.S.C. § 263a(f)(4)(B)(iv) (emphasis added). The final rule issued by HHS established a maximum daily work rate of no more than 100 actual patient slides in a 24-hour period, assuming an 8-hour workday—a workload of 12.5 patient slides per hour. 42 C.F.R. § 493.1257(b)(1). The proficiency testing protocol, on the other hand, requires cytologists to review only 10 slides⁷ in a 2-

⁷In the final rule, the agency explained that it was reducing the number of test slides from 20 to 10 in light of the fact that a 20-slide test would impose high testing costs and make it difficult to

hour period (a rate of 5 slides/hour). Between 30% and 60% of these slides must be abnormal, and to pass the technician must correctly identify at least 90% of the slides. 42 C.F.R. §§ 493.855(b), 493.945; 57 Fed. Reg. at 7041.

Consumer Federation argued that because the testing rate of 5 slides/hour is substantially less than the 12.5 slides/hour maximum permissible workload, the testing procedure does not conform to "normal working conditions." The district court agreed that the testing protocol violated CLIA's statutory mandate that testing take place, "to the extent practicable, under normal working conditions" and invalidated the regulation. We analyze this regulation under the same rubric as HHS's personnel qualifications regulation. If Congress has spoken clearly to the question of how the agency may conduct proficiency testing, we will give effect to Congress' intent. If not, we will sustain an agency reading of the statutory provision that is a "permissible construction" of CLIA.

Consumer Federation's challenge cannot be resolved under the first step of *Chevron* review. Although the statute requires HHS to make a reasonable effort to conform its testing protocol to actual working conditions, it does not require a precise replication of the workplace environment—witness the inclusion of the words "to the extent practicable" in the section. Nor has Congress defined with any precision when the agency may deviate from workplace conditions in the interests of practicality. We therefore turn to the second step of *Chevron*, and inquire whether the agency's interpretation of its statutory directive is a reasonable one.

The agency does not dispute that the testing rate it selected is significantly lower than the maximum daily work rate. The remaining questions, then, are (1) how much does the testing rate deviate from the normal working rate, if the average work rate is less than the maximum rate? and (2) do practical constraints on the administration of proficiency tests warrant such a deviation from normal working conditions? In its final rule, the agency explained the testing rate as follows:

We are modelling the scoring system as described in § 493.945 [the regulation implementing the proficiency testing program] after that in use in the State of Maryland. To that end we have changed the minimum passing score to 90 percent.... [W]e have added a maximum time allowed for each testing event, based on the PT

acquire enough high-quality slides to implement a national testing program. 57 Fed. Reg. at 7050.

program in the State of Maryland. Individuals are given not more than 2 hours to complete a 10-slide test and 4 hours to complete a 20-slide test. These time limits were established to provide for equitable testing on a national scale and to allow individuals sufficient time to complete the test at their normal pace without unduly restricting or extending the time for the examination.

57 Fed. Reg. at 7041. This explanation is simply too terse to support the agency's decision to use a testing rate which is less than half the maximum work rate, in the face of statutory language directing it to test under normal working conditions to the extent practicable. HHS first states that it has chosen this rate because it is identical to the rate used in the Maryland proficiency program. It provides no reason, however, as to why it has selected Maryland's program as a prototype, or why Maryland itself uses a testing rate of 5 slides/hour. The agency then claims that its rate allows for "equitable testing" and gives all cytologists "sufficient time" for the test. It is decidedly unclear exactly what the agency means by this. What is the average work rate, if it is less than the maximum rate of 10 slides/hour? Is it 5 slides/hour? Or do the slowest cytologists work at this rate? If so, why is it impractical for the agency to test at the average rate, even if some below-average cytologists fail the test as a result? Are there ways in which the demands of a proficiency test differ from normal working conditions—such as the argument proffered in the Collins declaration, discussed below, that a proficiency test includes more abnormal slides than are viewed in a regular workday—and if so, are such departures from normal conditions warranted? Without further definition of normal working conditions and an explanation of why the testing protocol it has selected conforms to these conditions, "to the extent practicable," we are at a loss to understand how HHS's proficiency testing regulations reflect a reasonable interpretation of the relevant CLIA provision.

HHS did belatedly attempt to remedy this problem. In a supplemental declaration submitted to the district court, an HHS official argued that the slower rate used in the testing protocol was necessitated by the higher percentage of abnormal slides included in the test. In an average workday, approximately 5% of the slides reviewed by a cytologist are abnormal. By contrast, between 30% and 60% of the slides in a proficiency test are abnormal, in order to test the cytologist's knowledge of a wide range of abnormalities. The official also stated that evaluation of an abnormal slide requires more time than examination of a normal slide because the cytologist must classify the specific type

of abnormality she sees. Since the testing protocol includes three to six times more abnormal—that is, more time-consuming—slides than occur on an average workday, it is not practical to require cytologists to review them at the normal work rate. See Collins Decl. of July 29, 1993, reprinted in App. at 68, 70-71.

Even if this rationale might have passed as sufficient grounds for the agency's decision to use a testing rate of 5 slides/hour, the critical fact is that HHS did not proffer it during the rulemaking process. Instead, the agency submitted the Collins declaration only at the district court stage. Our review of HHS's action can be based only on the administrative record, not "some new record made initially in the reviewing court." *Camp v. Pitts*, 411 U.S. 138, 142 (1973); see also *Center for Auto Safety*, 956 F.2d at 314 (rejecting agency's rationale for its bridge inspection regulations as *post hoc* rationalization not included in administrative record). As the government points out, this court has on occasion flexed the "record requirement" to allow the admission of agency declarations that "merely illuminate reasons obscured but implicit in the administrative record." *Clifford v. Pena*, 77 F.3d 1414, 1418 (D.C. Cir. 1996) (quoting *Appeal of Bolden*, 848 F.2d 201, 207 (D.C. Cir. 1988)). But the Collins declaration falls on the "*post hoc*" side of this administrative divide. Rather than simply providing additional background information about the agency's basic rationale for the proficiency testing protocol, as did the agency's declaration in *Clifford*, Collins' statement offers an entirely new theory for the testing rate selected by HHS—that a proficiency test should include a higher proportion of abnormal slides than are read in an average day, and that a slower-than-average work rate is appropriate because each of these abnormal slides takes more time to review than a normal slide. The only information along these lines included in the rulemaking record or statement of basis and purpose is a description of the criteria for scoring a proficiency test, from which it is possible to deduce the number of abnormal slides included in a 10 slide proficiency test. Without more, the mere fact that the test includes a higher proportion of abnormal slides than occurs in an average workday tells us little about the agency's choice of a testing rate. Since the agency has not up to now provided an adequate explanation on the record of why its testing protocol represents a permissible interpretation of the pertinent CLIA provision, we remand to the agency to articulate a

convincing rationale for its protocol or to continue the rulemaking process it has already commenced for issuing a new one.⁸

So ordered.

⁸We agree with the district court that this regulation may remain in place until a new explanation or a new rule is issued. HHS also challenges the district court's order to engage in expedited rulemaking. Such an order constitutes extraordinary relief, and is to be granted only upon a finding of unreasonable delay or imminent risk to public health and welfare. *See Telecommunications Research & Action Center v. FCC*, 750 F.2d 70, 77, 79-80 (D.C. Cir. Note 8—Continued 1984); *Public Citizen Health Research v. Commissioner, Food & Drug Admin.*, 740 F.2d 21, 32 (D.C. Cir. 1984); *Public Citizen Health Research Group v. Aughter*, 702 F.2d 1150, 1157-58 (D.C. Cir. 1983). The district court identified no evidence suggesting that the agency had engaged in unreasonable delay, and while the health risks posed by unreliable clinical tests are indeed serious, the court did not find that the existing CLIA regulations posed a "significant risk of grave danger." *Cf. Aughter*, 702 F.2d at 1157. Accordingly, the order to expedite rulemaking was inappropriate.

In light of our ruling upholding the agency's personnel qualifications regulation, the agency obviously need not take further action with respect to that rule. As to the cytology proficiency testing, the agency has already issued a new proposed rule. *See* 60 Fed. Reg. 61,509 (1995). Although our ruling requiring a remand for an adequate explanation of this rule supercedes the district court's mandate, if the agency chooses to issue a new rule instead of proffering an adequate explanation for the regulation already in place, it may resume the rulemaking process it has already initiated in response to the district court's order.