the program can drive improvements in emergency and disaster care for children.

Madam Speaker, I want to acknowledge the bipartisan nature in which this bill moved through our committee, working on both sides of the aisle within the Energy and Commerce Committee. We worked together to make this bill as good as it can be.

Madam Speaker, nobody likes to see a child get hurt. Together, we can assure that when that happens, children have the best possible chance for recovery and a good outcome. I strongly urge the adoption of this legislation.

Mr. DEAL of Georgia. Madam Speaker, I urge the adoption of this resolution.

I yield back the balance of my time.

Mrs. CAPPS. Madam Speaker, I am pleased to yield 3 minutes to the gentleman from North Dakota (Mr. POMEROY).

Mr. POMEROY. I thank the gentlelady for yielding, and I am also very pleased to speak in favor of H.R. 2464, the Wakefield Act.

I wanted to bring you just a little bit of perspective in terms of the difference this act has made in one young man's life, and I think it's reflective of a number of children who have been saved by having medical appropriate services for traumatic and life-threatening injuries of kids.

The Wakefield Act is called the Wakefield Act in recognition of a living memory of a family, the family of Tom Wakefield, who was involved in a horrible head-on traffic accident as they drove to the airport for a winter's vacation. A vehicle crossed the median and struck this vehicle head on, killing Tom and two of his children, one age three and one age seven. Twelve-year-old Lucas lost his arm in the accident and was almost lost as well.

Emergency responders on the scene and thereafter saved his life and the life of his mother, Loy. I know this family, and I know their survivors, and I care deeply about them. They have certainly impressed upon me, as they would impress upon any of you, just how vitally important it is that we equip our emergency response to deal with any who may be hurt. And the 40 percent improvement in saving lives of children since the act was initially passed in 1984 shows just how critically important this reauthorization is. I'm very pleased that the Commerce Committee has done the work to bring it to the floor today, and I am grateful for the chance to speak on the bill.

I was at an event just this weekend where Lucas, now fully recovering, adapted to his new circumstance. This is a young man that makes me very, very proud. And I believe the Wakefield Act, named in honor of his family, is a very appropriate commendation of the ongoing efforts to keep all our children safe.

Mrs. CAPPS. Madam Speaker, I have no further requests for time. And following that eloquent testimony to the value of this legislation, we can all recognize that H.R. 2464 is an important measure that will work toward ensuring the best emergency medical care for all children.

I again want to congratulate my colleague on the Energy and Commerce Committee, JIM MATHESON, and all of those who have spoken today, including the ranking member of the subcommittee, for all the hard work and dedication to this important piece of legislation. I urge all of my colleagues to join in support of H.R. 2464.

Mr. KING of New York. Madam Speaker, today I rise as a strong supporter of H.R. 2464, the Wakefield Act, which will reauthorize the Emergency Medical Services for Children program for an additional 4 years.

Since the program began in 1984, EMSC grants have helped all 50 States to better prepare their health systems to treat children in an emergency. The EMSC program has improved the availability of child-appropriate equipment in ambulances and emergency departments, supported hundreds of programs to prevent injuries, and provided thousands of hours of training to EMTs, paramedics, and other emergency medical care providers.

In my home State of New York, EMSC funds are going toward the development of a statewide, standardized system that recognizes hospitals capable of managing pediatric emergencies, both trauma and medical. This will enhance the State's ability to transfer injured children to the hospital best suited to their treatment. New York is also utilizing EMSC funds to ensure that all ambulances have the essential pediatric equipment and supplies for prehospital pediatric emergency care.

Across the country, EMSC is enabling State and local emergency care providers to better treat children. The projects funded under EMSC are vital for the safety and well-being of America's children and have saved countless lives throughout the program's existence. During a time when a terrorist attack or natural disaster may occur at any moment, it is essential that we ensure that we are adequately prepared to care for every infant, toddler, and child in an emergency situation.

I would like to thank Representative MATHESON for his hard work and continued leadership on this issue, and I urge you to support the Wakefield Act.

Mrs. CAPPS. Madam Speaker, I yield back the balance of my time.

The SPEAKER pro tempore. The question is on the motion offered by the gentlewoman from California (Mrs. CAPPS) that the House suspend the rules and pass the bill, H.R. 2464, as amended.

The question was taken.

The SPEAKER pro tempore. In the opinion of the Chair, two-thirds being in the affirmative, the ayes have it.

Mr. MATHESON. Madam Speaker, on that I demand the yeas and nays.

The yeas and nays were ordered.

The SPEAKER pro tempore. Pursuant to clause 8 of rule XX and the Chair's prior announcement, further proceedings on this motion will be postponed.

CYTOLOGY PROFICIENCY IMPROVEMENT ACT OF 2008

Mrs. CAPPS. Madam Speaker, I move that the House suspend the rules and pass the bill (H.R. 1237) to amend the Public Health Service Act to provide revised standards for quality assurance in screening and evaluation of gynecologic cytology preparations, and for other purposes, as amended.

The Clerk read the title of the bill.

The text of the bill is as follows:

H.R. 1237

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the ''Cytology Proficiency Improvement Act of 2008''.

SEC. 2. REVISED STANDARDS FOR QUALITY ASSURANCE IN SCREENING AND EVALUATION OF GYNECOLOGIC CYTOLOGY PREPARATIONS.

(a) IN GENERAL.—Section 353(f)(4)(B)(iv) of the Public Health Service Act (42 U.S.C. 263a(f)(4)(B)(iv)) is amended to read as follows:

''(iv) requirements that each clinical laboratory—

''(I) ensure that all individuals involved in screening and interpreting cytological preparations at the laboratory participate annually in a continuing medical education program in gynecologic cytology that—

''(aa) is approved by the Accrediting Council for Continuing Medical Education or the American Academy of Continuing Medical Education; and

''(bb) provides each individual participating in the program with gynecologic cytological preparations (in the form of referenced glass slides or equivalent technologies) designed to improve the locator, recognition, and interpretive skills of the individual;

''(II) maintain a record of the cytology continuing medical education program results for each individual involved in screening and interpreting cytological preparations at the laboratory;

''(III) provide that the laboratory director shall take into account such results and other performance metrics in reviewing the performance of individuals involved in screening and interpreting cytological preparations at the laboratory and, when necessary, identify needs for remedial training or a corrective action plan to improve skills; and

''(IV) submit the continuing education program results for each individual and, if appropriate, plans for corrective action or remedial training in a timely manner to the laboratory's accrediting organization for purposes of review and on-going monitoring by the accrediting organization, including reviews of the continuing medical education program results during on-site inspections of the laboratory.''.

(b) EFFECTIVE DATE AND IMPLEMENTATION; TERMINATION OF CURRENT PROGRAM OF INDIVIDUAL PROFICIENCY TESTING.—

(1) EFFECTIVE DATE AND IMPLEMENTATION.—Except as provided in paragraph (2), the amendment made by subsection (a) applies to gynecologic cytology services provided on or after the first day of the first calendar year beginning 1 year or more after the date of the enactment of this Act, and the Secretary of Health and Human Services (hereafter in this subsection referred to as the ''Secretary'') shall issue final regulations implementing such amendment not later than 270 days after such date of enactment.

(2) TERMINATION OF CURRENT INDIVIDUAL TESTING PROGRAM.—The Secretary of Health and Human Services shall terminate the individual proficiency testing program established pursuant to section 353(f)(4)(B)(iv) of the Public

Health Service Act (42 U.S.C. 263a(f)(4)(B)(iv)), as in effect on the day before the date of the enactment of subsection (a), at the end of the calendar year which includes the date of enactment of the amendment made by subsection (a).

The SPEAKER pro tempore. Pursuant to the rule, the gentlewoman from California (Mrs. CAPPS) and the gentleman from Georgia (Mr. DEAL) each will control 20 minutes.

The Chair recognizes the gentlewoman from California.

GENERAL LEAVE

Mrs. CAPPS. Madam Speaker, I ask unanimous consent that all Members may have 5 legislative days to revise and extend their remarks and include extraneous material on the bill under consideration.

The SPEAKER pro tempore. Is there objection to the request of the gentlewoman from California?

There was no objection.

Mrs. CAPPS. Madam Speaker, I yield myself such time as I may consume.

Madam Speaker, I rise in support of H.R. 1237, the Cytology Proficiency Improvement Act of 2007. This legislation would modernize Federal regulations under the Clinical Laboratory Improvement Amendments Act of 1988, CLIA, that subject those who screen and interpret Pap tests to annual proficiency testing.

In 2005, CMS launched a program to begin testing pathologists and other laboratory professionals who performed Pap tests for proficiency. However, the program was designed using regulations written in 1992. In the 13 years between the regulation and the program's start, significant investments were made in the science and practice of Pap tests. Instead of relying on outdated practices, H.R. 1237 draws on the best that science and technology has to offer.

H.R. 1237 has 175 bipartisan cosponsors, including myself and every other female member of the Energy and Commerce Committee. Additionally, this bill is supported by the College of American Pathologists, the American Medical Association, the American Clinical Laboratory Association, the American College of Obstetricians and Gynecologists, and the American College of Nurse Midwives.

I want to commend my colleagues, Representative GORDON and Representative DEAL, for their hard work and commitment on this very important piece of legislation. This bill would improve the quality of women's health care. I strongly encourage all of our colleagues to join me in support of H.R. 1237.

Madam Speaker, I reserve the balance of my time.

Mr. DEAL of Georgia. Madam Speaker, I yield myself such time as I may consume.

I, too, rise in support of the Cytology Proficiency Improvement Act. I was a sponsor of legislation similar to this in the last Congress which passed the House, but unfortunately it was never signed into law. The bill revises national quality assurance standards of laboratories responsible for cytology services.

A few summers ago, I had the opportunity to visit a laboratory of a pathologist in my district, and I saw first hand the impact of this legislation. This bill is the result of actions taken in 2005 by the Centers for Medicare and Medicaid Services to institute a proficiency testing program for individual pathologists.

□ 1530

Unfortunately, this program was based on regulations first issued in 1992 as a result of the Clinical Laboratory Improvement Amendments of 1988. Thus the cytology proficiency program is now very outdated and based on regulations from nearly 15 years ago.

The legislation would provide for an orderly phase-out of the current program and transition into a new program where all individuals involved in screening and interpreting Pap tests would participate in a continuing medical education program in gynecologic cytology. This educational approach will present participants with complex cases to keep their skills on the cutting edge and will provide individuals an opportunity to test their skills.

I believe this legislation would be an important step in the right direction and would modernize the current regulatory framework while providing quality assurance, as was required in the Clinical Laboratory Improvement Amendments. Unlike last Congress, I hope we will be able to get this legislation signed into law in order to modernize an outdated proficiency testing program for pathologists.

Madam Speaker, I reserve the balance of my time.

Mrs. CAPPS. Madam Speaker, I continue to reserve the balance of my time.

Mr. DEAL of Georgia. Madam Speaker, I am pleased to yield 5 minutes to my colleague from Georgia (Mr. PRICE), one of the original cosponsors of the legislation this year, a medical doctor.

Mr. PRICE of Georgia. I thank my friend and colleague from Georgia, Congressman DEAL, for his leadership on this issue and for the time today.

I also want to express my gratitude and thanks to Representative GORDON, who was extremely cooperative and helpful and productive throughout this entire process. I want to thank the American College of Pathology and all of the pathologists across the Nation who are working day in and day out to make certain that they provide quality care for the patients for whom they are charged.

Madam Speaker, I include in the RECORD a copy of an article by Dr. George Nagy that documents the dysfunctional federally mandated proficiency test in cytopathology.

THE DYSFUNCTIONAL FEDERALLY MANDATED PROFICIENCY TEST IN CYTOPATHOLOGY—A STATISTICAL ANALYSIS

Proficiency testing in cytopathology and in other disciplines should be based on firm statistical and scientific foundations, because test theory in general is a heavily statistical subject. Statistical considerations have demonstrated that the design of ''short'' proficiency tests in cytopathology, including the current federally mandated test, fundamentally is unsound because of the lack of sufficient validity and reliability. Examinees too frequently are misclassified by such short-format tests: Competent examinees fail the test in surprisingly high numbers, whereas most of the examinees who have insufficient cytologic skills eventually pass the test after the allowed retakes. Only dichotomous tests are suitable for accurate computation of the effects of test design on reliability, but the statistical conclusions also are generalizable to nondichotomous tests. In conclusion, the current federally mandated proficiency test cannot reliably measure the level of expertise of cytologists and, thus, cannot assure that only adequately skilled individuals evaluate Papanicolaou test samples. To render the test suitable for its intended purpose, the authors believe that complete redesign of the test, with the participation of experts in modern test theory, would be advisable.

Proficiency testing in cytopathology (PTC), which was established in the 1991 regulations to implement the Clinical Laboratory Improvement Amendments of 1988 (CLIA'88), has only recently been enforced on a national scale. For more than a decade, during which logistical hurdles hampered the development of a national program for PTC, there was not much incentive to think about the value and potential of PTC or its theoretical background or to worry that the test design was so poor. In 2004, however, the Center for Medicare and Medicaid Services announced that a national PTC program developed by the Midwest Institute for Medical Education had been approved and that the regulations finally would be enforced on a national level. Suddenly, the shortcomings of the test were everyone's problem. What followed was a flurry of comments, articles, proposals, and Internet discussions about the PTC and its future. Although the testing has proceeded nationwide in conformity with the original regulations, the dust has not yet settled on the subject. The professional organizations agree that PTC, as prescribed in CLIA'88, is inadequate and is in great need of improvement if indeed it should remain in place at all. Regarding the projected revisions, it is a real impediment that some regulatory authorities that are in a position to make decisions about the implementation of PTC apparently are not familiar with most of the theoretical implications of test theory, which is an exceedingly complicated subject. So long as the test is mandatory for every practitioner of gynecologic cytopathology in the United States, it is in the best interest of all participants for PTC to become a scientifically well-founded, valid, and reliable quality assurance method. In the current article, we have attempted to shed light on some gaps in the knowledge about the theoretical underpinnings of PTC that seem to endure in the cytopathology literature.

TEST THEORY IS STATISTICAL

Test theory is a heavily statistical subject. Virtually all aspects of test theory have been investigated in depth almost exclusively by educators and psychologists, which is understandable, because testing is a central issue in their disciplines. Unfortunately, this valuable body of literature apparently has been disregarded completely by the federal authorities that are responsible for PTC regulations.

The statistical apparatus used in modern test theory is formidable. Many books and

articles written about the subject use highly sophisticated mathematical tools, including differential and integral calculus and matrix algebra. One of the reasons for the high degree of mathematization of test theory in psychology and education science is that these disciplines deal largely with intangibles, like motivation, intelligence, understanding, and adaptability, which are not directly measurable. Such entities must be studied indirectly, through measurements of other quantities. That is why psychological test theory introduced the concept of ''constructs'' that can substitute for and represent the kinds of abstract attributes mentioned above. Even so, the highly complicated mathematical and statistical tools that have been promoted in educational and psychological test theory fulfill mainly academic purposes. Most actual problems in everyday testing can be solved on a practical level that does not use highly complicated mathematical methods but, at the same time, does not disregard basic statistical principles.

TESTING IN THE PHYSICAL AND BIOLOGIC SCIENCES

Cytopathology, unlike educational science or psychology, is an applied natural science, and this is one of the reasons why PTC can be performed without the application of overly sophisticated mathematical tools. Interpretation of Papanicolaou smears, reproduction of cytologic diagnoses, and measurement of false-negative proportions, among others, are very complex tasks. By comparison, technically, it is a comparatively straightforward matter to evaluate the examinees' ability to assign diagnostic categories to cytologic changes observed on a slide or computer screen. Thus, abstract constructs hardly are needed in PTC. Nevertheless, a certain level of mathematical and statistical understanding by the designers of the test is crucial if a fair and scientifically valid system of PTC is to be established. Most pathologists, including ourselves, do not have rigorous training in statistics; therefore, if PTC is to continue, then the regulatory authorities ought to contract with experts in statistics and test theory who, through interaction with knowledgeable cytopathologists and cytotechnologists, would design an equitable and scientifically well-founded system for the nationwide PTC.

We do not mean to suggest that statisticians have not participated in the design of cytology testing programs. In fact, the College of American Pathologists' (CAP) Interlaboratory Comparison Program for Cervicovaginal Cytology was designed, implemented, and monitored with the extensive help of statistical expertise. However, this educational endeavor was not intended to be a PTC program as envisioned in the federal regulations. In fact, its original, scientifically and statistically supported structure ironically prevented its use as a PTC program because of the specific requirements of the federal regulations.

SHORT TESTS AND RELIABILITY

One of the central problems in the practice of PTC is reliability, and the reliability of PTC is related closely to the size of the test sets (the number of the test items or challenges in 1 test set). ''Short'' tests, which require the evaluation of relatively small numbers of slides, are characterized by a high misclassification rate. (The pervasive effect of sample size on the reliability of statistical inference is the reason why pollsters use large samples: The larger the sample, the narrower are the confidence limits in relative terms. The statistical estimates inferred from a single sizable sample that has been chosen by randomization will approach the true parameters of the population.) Short tests will not prevent the frequent failure of competent examinees or the passing of examinees who have less than desirable skill levels. Already in 1991 one of us (G.K.N.), in a report that was written with D.C. Collins, emphasized that the expected misclassification rate of such short tests can be surprisingly high and that, in the case of dichotomous tests, this rate can be calculated (or approximated) through the use of the binomial theory of statistics. (A dichotomous test evaluates the responses to test items as ''right'' or ''wrong,'' without using intermediate results or weighing of answers. The PTC system used in New York State for 36 years was dichotomous and so was the original Interlaboratory Comparison Program in Cervicovaginal Cytology. The CLIA'88-mandated PTC is not dichotomous.) This so-called ''simple binomial error model'' was described in test theory initially in the 1950s.

The results of the CLIA'88 mandated national PTC in 2005 dramatically demonstrated the effect of misclassification during short tests, as described previously. According to the data from the National Cytology Proficiency Testing Update, 9% of the examinees failed the test when they attempted it for the first time. However, when this group that supposedly had inferior skills retook the test, curiously, the failure rate for this second attempt was similar to that for the entire original group (10%). It appears that the cytologic skills among those examinees who had failed originally improved miraculously, allowing 90% of them to pass the examination, although all of them initially failed. It is hard to believe that a short remedial training between the first and second attempt could result in such an impressive real improvement. The only plausible scientific explanation is the well-known statistical phenomenon, the Galtonian ''regression toward the mean.'' The majority of failures during the first attempt were the consequence of misclassification because of the poor validity and reliability of the short test and were not caused by the insufficient skills of those who failed. The failure rate in all groups of examinees is about the same on the first attempt and on the second attempt, and previous failures do not seem to matter much. Essentially, the results of the CLIA'88-mandated PTC mostly mirror the statistical chances and not the examinees' skills.

Of course, multiple other variables beyond regression toward the mean, including experience gained in the technique of the test, differences in the difficulty of particular test sets, and even increased skills after remedial training, etc, also may play a role in the improvement of test results at the second attempt for individual examinees. However, to date, we do not have any data or even a plausible explanation concerning how any of these other factors, with the exception of regression toward the mean, could produce such a consistent result.

THE SIMPLE BINOMIAL ERROR MODEL

Misclassification of examinees by any short test, including the CLIA'88-mandated PTC, can be demonstrated by means of an analogy. Strictly speaking, this analogy is applicable only to dichotomous testing systems. However, in this sense, dichotomous and non dichotomous systems are correspondent. For statistical or evaluation purposes, non dichotomous systems can be made dichotomous at any time, even after the tests have been carried out. For example, an answer can be evaluated as correct only if it falls into the appropriate single category (''success'') and all other answers are rated as wrong (''failure''). Another solution to this problem in PTC would be to restrict the number of diagnostic categories to 2, with 1 category, for instance, ''negative for premalignant or malignant changes'' and the second category ''premalignant or malignant lesions are present.'' This is the approach used in the original CAP PAP program with its ''100 series'' and ''200 series.''

The CLIA'88 regulations concerning PTC, with their 4 diagnostic categories and complicated scoring system, do not fit into the dichotomous scheme. Despite this fact, the conclusions drawn by using the binomial error model regarding PTC are applicable to any short test to a large extent.

EXAMPLE OF SIMPLE BINOMIAL ERROR MODEL

For the purpose of illustration, let us suppose, that in a large population (for instance, that of an entire country), the results from a scrupulous statistical survey using many thousands of questionnaires and proper randomization indicate that the proportion of individuals who like to watch television (TV) is 90%. Because the survey is conducted in a scientific way and the sample size is very large, this result is considered highly accurate. The basic question on which the analogy with PTC will be based is, ''What can we expect if we ask 10 randomly selected individuals in this population about their attitude toward TV?'' The most probable result will be that, in this population, 9 of 10 individuals will like TV. However, it is reasonable to expect that, in many samples that consist of 10 individuals, all 10 individuals are TV fans; whereas, in other similar samples, there may be only 8, 7, or 6 such individuals. However, it is hardly conceivable that we will identify as few as only 1 or 2 fans in a sample of 10 individuals if the principle of random selection is followed.

Random selection is important. For example, a nonrandom sample, like one that consists exclusively of nuns in convents, would not yield a statistically valid reflection of the entire population; indeed, we may identify only 1 or 2 individuals in such a sample who like to watch TV. Exclusive selection of nuns or members of any other group with some special interest would not be compatible with the principle of randomness. However, to select a nun occasionally in a sample, with a frequency roughly corresponding to the proportion of nuns in the entire population, would be appropriate.

There is a statistical method that uses the so-called ''binomial formula'' for calculating the probability of encountering 10, 9, 8, 7, etc, TV fans in a sample of 10 individuals from our postulated population. (This method is not detailed in the current article, but an explanation can be found in any elementary statistical textbook). The probabilities even can be looked up in tables that are found at the end of statistical books. Under the circumstances outlined above (with a 90% proportion of TV fans in a sample size of 10 individuals). the probabilities of identifying 10, 9, 8, 7, and 6 TV fans in a random sample of 10 individuals are 0.35, 0.39, 0.19, 0.06, and 0.01, respectively.

The probability of identifying ≤5 TV fans under the above-described circumstances in a truly random sample of 10 individuals is exceedingly small. The succession of numbers described above represents a ''probability distribution,'' which can be observed in a histogram. This distribution is interpreted as follows: If, from this very large population, we take numerous random samples, each consisting of 10 individuals, and ask about their preferences for TV; then we will find that 35% of the samples would include 10 fans, 39% of the samples would include 9 fans, 19% of the samples would include 8 fans, and so on.

If we change the size of the sample, then the magnitudes of the single probabilities

and their distribution also will change and, along with them, the probability distribution. If we choose sample sizes of 100 individuals instead of 10, then the probabilities will be clustered much more tightly around the value of 90% than was the case in the smaller samples. The larger the size of the sample, the more reliable is the estimation; in other words, the observed value in every sample approaches the real population parameter. It is virtually unimaginable that there will be only 50 or 60 TV fans among 100 randomly selected individuals from this population. (Distribution data for such large samples are not provided even in the tables of larger statistical reference books: They are not needed, because the probability distribution for large samples can be found by the so-called ''normal approximation of the binomial distribution.'' To perform this method is mathematically simple, but the results may be slightly inaccurate. There are complex Web-based Internet tools, however, that calculate these probabilities very accurately.) Of course this holds true only if the randomness principle is strictly observed.

How can we apply the reasoning described above to the issue of sample sizes in PTC? Fortunately, the results of these binomial calculations can be generalized. The reason why we can do this is that, if the ''experiment'' qualifies as binomial, then the specifics of the experiment, whether they are related to liking TV or to success in PTC, have no bearing on the values of the probabilities or on the probability distribution.

TRUE SCORES

At this point, we need to review the term ''true score,'' a concept that is used widely in modern test theory. The true score of a hypothetical examinee is defined as the average of the observed or measured scores that would be obtained over an infinite number of repeated testing by the same test, provided that the examinee's skills remain indefinitely stable. For actual examinees, the true score can be estimated with a small error margin, but its exact value is essentially unknowable. For instance, if a cytologist screens 100,000 cervical smears, and if his or her diagnoses are correct 98,000 times, then the approximation of his or her true score is 0.98. Because the accurate determination of the true score would require an infinite number of repeat testing, which is not feasible, this true score of 0.98 remains an approximation. Obviously, we can be rather sure that, when the same individual screens the next 100,000 preparations, the approximation of his or her true score will not remain the same: The chances of this are infinitesimally small. The estimate of the true score will almost certainly change slightly, for instance to 0.97 or to 0.99, and so on, for each successive trial.

It has to be emphasized that assignment of an exact ''true score'' to a cytologist is somewhat arbitrary for further reasons. It cannot be expected that anybody's cytologic skills will remain invariant for a prolonged time. We can hope, of course, that the professional prowess of cytologists improves over time. Furthermore, everybody who has ever screened cytology specimens knows that screening performance depends on many factors, some of which are extraneous to the level of cytology skills. On a ''good'' day, a cytologist may function on a 0.98 score level; whereas, on a different, ''bad'' day, he or she might be less ''proficient.'' Even his or her experience with particular kinds of cytologic presentations on the previous day, for example, having seen an unusual presentation of high-grade squamous intraepithelial lesion on a quality-assurance review, could affect decision-making on the current day. Of course, these and other psychological variables (eg, the effects of anxiety or tiredness during tests or routine work) cannot be factored into the statistical considerations. Nagy and Collins, describing this concept, used the term ''competence level'' instead of ''true score'' in their 1991 article.

Direct measurement of the true score is not possible. What we have after an evaluation of test results is the ''observed score,'' which is related to the true score but is not identical to it. It can be considered an estimate of the true score.

COMPARISON OF TV PREFERENCE AND PTC RESULTS

TV preference and PTC results can be compared as follows: The values derived by the binomial formula are determined only by the number of trials and the probability of success. If the ''experiment'' qualifies as binomial, then the specifics of the experiment have no bearing on the numerical results. (In statistical parlance, any methods or procedures that yield raw data are called experiments.) In our TV example, the number of trials (the sample size) is 10, and the probability of success is 0.9. These 2 data are sufficient to calculate the probability distribution for this specific case. Let us consider now an example of PTC in which these specifics are the same as described above. The PTC design prescribes 10 slide test sets (number of trials). A cytologist who performs routine screening and customarily renders accurate diagnoses 9000 times among 10,000 screened slides has an approximate true score of 0.9. (In other words, the probability of success is 0.9.) When this cytologist attempts to pass this particular PTC, then the probability distribution of the possible correct answers will be identical to the probability distribution observed in the TV example, because the specifics of the TV experiments are the same. If this hypothetical cytologist attempts the test many times, then he or she will read 10 slides correctly in 35% of the tests, 9 slides correctly in 39% of the tests, and so on. The numerical values in the 2 experiments are identical.

We also should note that, if an examinee reads 10 slides or 9 slides correctly:which happens in 74% of events under the circumstances described above, then he or she passes the test. However, this individual, who essentially has an adequate true score, will fail a dichotomous PTC 26% of the time because of the low validity and reliability of the test. The phenomenon of failure in this case can be called ''type 1 error.'' (The null hypothesis is that ''the cytoscreener is competent.'') A valid and reliable test is expected to pass virtually all cytoscreeners with true scores on the 0.9 level; however, any dichotomous test that consists of 10 slides or challenges will misclassify approximately 26% of such individuals. It is obvious that this test does not really meet the expectation to determine the competence of an examinee who had a true score of 0.9.

It needs to be reiterated here that binomial calculations can be performed only for dichotomous tests. The probabilities for some well ordered, nondichotomous tests may be calculated by the use of more complicated multinomial assessments.

LIMITATIONS OF THE SIMPLE BINOMIAL ERROR MODEL

The binomial error model provides only a rough appraisal of the statistical factors that need to be taken into account in the design of PTC. One of the drawbacks of the model, as mentioned above, is that it is applicable only to dichotomous testing systems. However, the simplicity, transparency, and mathematical calculability of dichotomous setups counterbalance every other consideration. The dichotomous test design makes it possible to assess the impact of test set size on test validity and reliability and to calculate confidence intervals. Thus, the use of a dichotomous test would confer greater predictability and practicability to PTC. The effects on test validity and reliability of a haphazard design, like the CLIA'88-mandated PTC, hardly are calculable by scientific-statistical means. We do not state that dichotomous designs would solve every problem inherent in every type of test, including PTC. However, given that all other conditions of the testing are equal, dichotomous tests have insurmountable advantages over nondichotomous tests.

SIZE OF TEST SETS AND RATE OF MISCLASSIFICATION

Figures (not shown) illustrate the probability distributions of correct diagnoses for variable test set sizes and for examinees with different theoretical ''true scores.'' An ideal and flawless PTC would fail all examinees with true scores of 0.85, but no test design can fulfill such requirements. The reliability of the tests improves, however, as the test sets get larger. For examinees with true scores of 0.85 or 0.8, the accuracy of the test increases in parallel with the increasing size of the test sets. (The failure rates become larger for larger test sets.)

Visualization of the effect of sample size on misclassification also is possible by tabulation. The more slides the test set contains, the lower the misclassification rate. There appear to be anomalies at the set sizes of 9 and 19, in which the misclassification rate decreases for examinees with low true scores and increases for the more competent examinees. A test set that consists of 9 or 19 slides would be a very impractical choice. If the passing level is set at 90% (eg, 9 correct answers for 10 slides in dichotomous tests), as it is the general practice for PTCs, then 1 error is allowed for a 10-slide set. Under these circumstances, to pass a test based on 9-slide sets with a 90% passing grade would be incomparably more difficult than to pass a test based on a 10-slide set, because a single mistake would mean an error >10% and, consequently, a failure. The situation is similar for 19- or 29-slide sets. The greater grade of difficulty with a 9-slide test set is reflected in the smaller passing rates for both competent and less competent examinees. (This circumstance, paradoxically, improves the accuracy of the test for the participants with low true scores.) For these reasons, if the passing level is set at 90%, then only decimal-based test set sizes (10, 20, 30, etc. slides or challenges) should be used.

Another observable phenomenon is the ''law of diminishing returns,'' in which, as the number of slides in the test sets is increases, the misclassification rates decrease. However, the rate of decrease is not level but trails off with increasingly larger set sizes. For instance, misclassification of examinees with a true score of 0.8 is almost halved, from 38% to 20%, when the number of slides in the sets increases from 10 to 20. The next step, from a 20-slide set to a 30-slide set, is accompanied by a smaller relative improvement, and so on.

An important conclusion that can be drawn is that, when the number of slides is increased in the test sets, the decrease in the misclassification rate is more precipitous if the true score is 0.8 or 0.85, ie, on the side of the table for less competent examinees, than if the true score is 0.95. From our viewpoint, this is an advantage. The basic purpose of PTC is not the confirmation of the proficiency of the average cytologist who performs well but the identification of individuals who may have problems with expertise and need remediation. The type 1 error, the failure of competent examinees, is less consequential than the type 2 error, the passing

of less competent examinees. The simple binomial model is more suitable to investigate the latter than the former in the set-size ranges that are prevalent in the practice of PTC.

WHAT SHOULD BE THE MINIMAL NUMBER OF TEST SLIDES IN TEST SETS?

The question about the minimal number of test slides in test sets could be formulated more accurately as follows: What should be the minimal number of test slides so that we can be 90% confident that the test result is accurate? This type of calculation is relatively simple to perform if the test is dichotomous. In our calculations, we assumed a dichotomous test and 90% as the passing level for the observed score.

The minimum necessary number of test slides depends to a large extent on the competence of the individual examinee. For a cytologist with very poor skills, a relatively small test set would suffice. However, the discriminatory power of PTC decreases at the point where the skills of the examinee are almost satisfactory but still insufficient. Therefore, for such an individual, the test sets should be much larger if we want 90% confidence. It would be unrealistic to expect any test to differentiate easily between an ''incompetent'' cytologist whose true score is 0.89 and a ''competent'' cytologist with a true score of 0.9.

Just to illustrate a possible solution, we calculated the minimal size of test sets for examinees who had a true score of 0.8. We wanted to have 90% confidence in the accuracy of the test result. (This means that at least 90% of examinees with a true score of 0.8 will fail the test if the test set contains the calculated number of test slides.) Similar calculations were performed for examinees who had a true score of 0.85.

For the calculation, we used the algorithm written by the Vassar Education Department, which is in the public domain and may be found on the Internet. According to the results, a 40–slide set would provide >90% confidence (exactly, 92.409% confidence) in the accuracy of the results for examinees with a true score of 0.8. A 30–slide set would provide only an 87.729% confidence level for these individuals.

For examinees with a true score of 0.85, much larger test sets would be necessary to provide 90% confidence in the results. A test set consisting of 90 slides would provide 88.468% confidence, and only the use of a 100–slide test set would ensure >90% confidence (exactly, 90.055 confidence) in the test results. The extent of the confidence intervals can be easily visualized. Lord et al. presented the 90% confidence intervals for a 30–item dichotomous test on different true score levels.

The numbers provided above are given only for illustrative purposes. It is obvious that test sets consisting of 100 slides, or even 40 slides, could not be used under the generally accepted conditions of PTC. Evidently, only a board-type, full-day, or 2-day-long examination would satisfy the statistical requirements for an accurate and equitable test. Conversely, because such a board-type test would determine the capabilities of the examinees with a high level of accuracy, it would become safe to increase the intertest interval to 8 years or 10 years.

However, if most aspects of the current federal regulations for PTC remain in force—in other words, if a highly inaccurate and unreliable test also will be used in the future—then it will not be advisable to increase the yearly interval between tests very much. The main reason for this is that short tests are incapable of accurately identifying examinees with low professional skills. Competent examinees who fail the test (type 1 error) pass the test on the second or third attempt with a high probability. Most of these valuable professionals are not harmed much beyond the inconvenience of repeated testing. In contrast, examinees with questionable skills who pass the test (type 2 error) do not have to submit to repeat testing, and they continue to screen patient slides without censure at least until the next test. Of course, it may be argued that, if the test were totally useless, then increasing the interval between test events would not have any effect on public health. However, if the test were totally useless, then the only honest course to follow would be the complete abolishment of PTC. In our opinion, the test in its present form is not totally useless. The current test will force a certain number of cytologists with very poor professional skills (regardless of their low proportion in the entire cytopathology community) to recognize their deficiencies, to participate in remediation(s), and at least to attempt to improve their professional skills. However, as made obvious in the discussion above, the federally mandated PTC in its current form is not able to identify all cytologists with very poor skills. Allowing such individuals, unidentified by the test, to continue screening constitutes a certain danger for the public. If we try to make the current PTC useful at least to some degree, then we should not increase the time interval between tests to 3 or 4 years.

THE HIGH PASSING RATE OF LESS SKILLED PROFESSIONALS IN SHORT TESTS

Through the use of the simple binomial model, it also is possible to calculate the number of less than competent individuals who eventually will pass the short tests after repeated attempts. For instance, among 100 examinees who have true scores in the less competent range of 0.85, 54 individuals will pass a dichotomous test that consists of 10 test slides on the first attempt. The remaining 46 examinees will attempt the test a second time, and 54% of them (ie, 25 individuals) will pass on this second try. The remaining 21 examinees will attempt the test a third time, and 54% of them (ie, 11 individuals) will pass. In summary, 54 + 25 + 11 = 90 of these less-skilled examinees among 100 who were supposed to be identified by the system will avoid serious consequences if a short, 10–slide-based dichotomous test with 3 permitted retakes is used.

A similar calculation illustrates that, among 100 examinees with true scores of 0.8, 76 individuals eventually will pass, if 3 attempts are allowed, in a 10 slide-set, dichotomous PTC system.

These numbers indicate all too clearly the utter uselessness of short dichotomous PTCs in terms of capability to identify less skilled cytologists. However, we do not go so far as to declare that short PTC systems, dichotomous or nondichotomous, are totally lacking in utility. Even a short test generates interest, creates opportunity for self-assessment, and possibly highlights deficiencies in some areas in the professional knowledge of the individual cytologist. This effect should be perceived as beneficial. Our personal experience indicates that very short educational tests, although they may not be suitable in themselves as statistical assessments of professional knowledge of individuals, almost always provide a welcome impetus for continuing education. A short PTC, as an educational experience, may remain a valuable quality-assurance method, although it is limited in scope. In this regard, other valuable educational activities, such as the CAP Pap program, have their full justification. However, we in the cytopathology community should persevere in our attempts to prevent the deleterious situation in which PTC remains an expensive and rather meaningless ritual; a test that, on repeated attempts, can be passed by virtually all competent cytologists, as expected, and also by a very high percentage of those who would be adjudged incompetent if a more reliable testing process were available.

STATISTICS ARE NOT EVERYTHING

A more intensive integration of statistical principles would be needed to make the current design of PTC more functional. However, we do not believe that, even if statistical principles were applied optimally to PTC, all of the inherent problems of testing could be eliminated. There are many nonstatistical facets of all tests, including PTC. For instance, because, in cytopathology, we are confronted with the morphologic manifestations of extremely complicated biologic systems, total equivalence in the difficulty of test challenges (that is, absolute conformity of corresponding slides in different test sets) cannot be achieved. Perhaps this can be overcome with computerized digital tests to some extent in the future.

LESSONS FROM THE SIMPLE MODEL OF DICHOTOMOUS PTC THAT CAN BE APPLIED TO THE DYSFUNCTIONAL FEDERAL DESIGN

We emphasize once more that the discussions and calculations above are based on the relatively simple model of dichotomous proficiency testing. The current CLIA'88–mandated test, with its elaborate scoring system and multiple diagnostic categories, is much more complicated; therefore, our conclusions cannot be transferred to it in any straightforward or easy way. The proportions of expected misclassification rates, the widths of confidence intervals, and other statistical parameters in nondichotomous systems cannot be calculated accurately by using the simple binomial model. In other words, the generalizability (''external validity'') of the foregoing statistical considerations to nondichotomous systems could be questioned. The Galtonian regression toward the mean in the results of the first year of the CLIA'88–mandated test, however, provides indirect evidence that misclassification by the federal test is substantial, and its magnitude is in the range indicated by the simple binomial model. Therefore, it is plausible that the conclusions of the statistical considerations outlined above are applicable to the federally mandated PTC to a large extent.

We emphasize that the theoretical underpinnings of PTC are much more complex than may be perceived readily. We hope that, if mandatory, nationwide PTC remains in any form, then it is redesigned to be a valid and reliable proficiency testing system or possibly a board-type examination. We believe that accomplishing this would require the engagement of both cytologists and experts who are well versed in the practical and theoretical aspects of modern test theory. This does not mean that more descriptive data from the existing results of the CLIA'88–mandated PTC should be collected. On the contrary, because the design of the CLIA'88–mandated test is flawed, little true insight may be gained by amassing and further studying descriptive data from such a source. Rather, we advocate the careful application of more inferential or theoretical statistics, which would allow a fairer conceptual design of PTC while leaving the final decisions in the hands of expert cytopathologists and cytotechnologists who are familiar the wider aspects of our difficult discipline.

I also want to thank all of the members of the Women's Caucus. Without their wonderful support, I don't know where we would be at this point. And I thank, once again, Congressman DEAL,

the ranking member of the subcommittee; Chairman PALLONE and Chairman DINGELL and Ranking Member BARTON.

Madam Speaker, as has been described by my colleagues, in 1998 the CLIA, or the Clinical Laboratory Improvement Amendments, went into effect. The law was passed. And it took them 4 years for the provision to evaluate the performance of laboratories interpreting Pap tests or Pap smears to be put into law or to have the rule finalized by Health and Human Services. The problem is that program then sat on the shelf for 13 years. So in 2005 the rules were then put into effect and enforced. And therein lies the program.

This program currently in place is based upon more than a decade old, even 15, 16 years old, 1992, regulatory approach that doesn't reflect the modern science and real-world laboratory practice. It does little to help patients or physicians charged with caring for them. The approach of relying on government-driven individual proficiency testing to evaluate the quality of Pap smear interpretations is both outdated and not cost effective.

So the solution is within the bill that we have before us today, H.R. 1237. There's a companion bill, Madam Speaker, over in the Senate, S. 2510, and I'm hopeful, as Congressman DEAL said, that we will be able to get this legislation through both Chambers during this session.

The Cytology Proficiency Improvement Act modifies CLIA by suspending the current regulation that subjects pathologists and others who screen for cervical cancer to annual proficiency testing and instead requires annual continuing medical education that would provide laboratory professionals opportunities to improve their screening and interpretation skills in a nonpunitive environment. The bill allows for an orderly phase-out of the current program and establishes reasonable timelines for the implementation of the new program. The educational approach is consistent with that included in the Mammography Quality Standards Act, a program that is remarkably effective. So the bill would ensure continuing education keeps up with the technology in the field and that clinicians are using day after day after day to help save lives of Americans all across our Nation. This is a major move in the right direction.

I want to thank once again all of those involved and encourage my colleagues to support the bill.

Mrs. CAPPS. Madam Speaker, I continue to reserve the balance of my time.

Mr. DEAL of Georgia. Madam Speaker, I urge the adoption of the bill.

Madam Speaker, I yield back the balance of my time.

Mrs. CAPPS. Madam Speaker, I have no further requests for time and again would like to commend my colleagues Representative GORDON and Representative DEAL and also the Women's Caucus for their much hard work and commitment on this important piece of legislation.

This bill would improve the quality of women's health care, and I strongly encourage all of our colleagues to join in support of H.R. 1237.

Mrs. MYRICK. Madam Speaker, I rise today in support of H.R. 1237, the Cytology Proficiency Improvement Act. I am pleased to see that the House will vote today on revamping a 16-year-old CMS regulation—from 1992—that calls for a Federal program to test the proficiency of individual laboratory professionals who read Pap tests.

I first became aware of the need to revisit this outdated regulation several years ago, in 2005, when CMS first began implementation of the program long after it was first put on the books. Congress knows well that promulgating regulations and implementation can do more harm than good.

The current oversight model that CMS is using is intended to help ensure that Pap tests are being read accurately—to improve public health. However, the approach established more than a decade ago, and being used today, doesn't necessarily protect women, improve quality or further our fight against cervical cancer.

H.R. 1237 provides an alternative. It redirects the current ''testing'' scheme to require pathologists and other lab technicians who read Pap tests to participate in an annual continuing medical education, CME program where their skills would be assessed and where the latest advances in Pap test practice could be shared. It would complement extensive Pap test quality controls that labs must already meet under the Clinical Laboratory Improvement Act. The Mammography Quality Standards Act includes a similar CME approach.

I've talked to pathologists in my district to better understand what it would take to add value to their profession, rather than just more red tape. Dr. Jared Schwartz was one of those who educated me and lent his expertise. He is now serving as president of the College of American Pathologists and is a strong advocate for ensuring access to Pap tests for all women. The laboratory and medical community support this bill, and I'm pleased to support it.

Mr. BUCHANAN. Madam Speaker, I rise today in support of H.R. 1237, the Cytology Proficiency Improvement Act of 2007. I am a cosponsor of this important legislation, which enhances women's health by establishing a continuing medical education requirement for pathologists and laboratory professionals who examine Pap tests to screen for cervical cancer.

I recently toured Sarasota Pathology and heard directly from my constituents about the importance of this bill and its potential to help save lives.

This legislation amends the Clinical Laboratory Improvements Amendments of 1988, CLIA, which mandated a cytology proficiency test to be administered by the Federal Government. However, the program lay inactive until 2005, which, because of scientific advancements makes the test obsolete and out of date.

Unlike the current CLIA testing model, H.R. 1237, with its annual continuing medical education requirement, will provide the means to increase the skills necessary to identify potential cervical cancer, and will keep pace with new science.

H.R. 1237 is modeled after the Mammography Quality Standards Act, MQSA, which was passed in 1992. That bill ensured women would have access to quality mammography procedures. This bill requires similar educational testing for pathologists.

The American Medical Association, the College of OBGYNs, the College of American Pathologists, the American Society for Clinical Pathology, the College of Nurse Midwifes, and the Cancer Research and Prevention Foundation endorse the bill.

Finally, I want to mention that the Congressional Budget Office has determined that it will not cost the Federal Government any additional expenditure.

Madam Speaker, I urge my colleagues to join with me in support of a bill that will greatly improve the quality of women's health care in America.

Mrs. CAPPS. Madam Speaker, I yield back the balance of my time.

The SPEAKER pro tempore. The question is on the motion offered by the gentlewoman from California (Mrs. CAPPS) that the House suspend the rules and pass the bill, H.R. 1237, as amended.

The question was taken; and (two-thirds being in the affirmative) the rules were suspended and the bill, as amended, was passed.

A motion to reconsider was laid on the table.

SAFETY OF SENIORS ACT OF 2007

Mrs. CAPPS. Madam Speaker, I move to suspend the rules and pass the Senate bill (S. 845) to direct the Secretary of Health and Human Services to expand and intensify programs with respect to research and related activities concerning elder falls.

The Clerk read the title of the Senate bill.

The text of the Senate bill is as follows:

S. 845

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the ''Safety of Seniors Act of 2007''.

SEC. 2. AMENDMENTS TO THE PUBLIC HEALTH SERVICE ACT.

Part J of title III of the Public Health Service Act (42 U.S.C. 280b et seq.) is amended—

(1) by redesignating section 393B (as added by section 1401 of Public Law 106–386) as section 393C and transferring such section so that it appears after section 393B (as added by section 1301 of Public Law 106–310); and

(2) by inserting after section 393C (as redesignated by paragraph (1)) the following:

''SEC. 393D. PREVENTION OF FALLS AMONG OLDER ADULTS.

''(a) PUBLIC EDUCATION.—The Secretary may—

''(1) oversee and support a national education campaign to be carried out by a nonprofit organization with experience in designing and implementing national injury prevention programs, that is directed principally to older adults, their families, and