Again, we would like to applaud the authors on their thoughtful and innovative curriculum design but encourage them and other readers to not adopt testicular self-examination as a screening tool for the detection of testicular cancer. We would also like to encourage primary care researchers to be mindful of research priorities so that the issues that we highlighted above would be less common.

Dean A. Seehusen, MD
John A. Edwards, MD
Department of Family Medicine
Madigan Army Medical Center
Tacoma, Wash

Acknowledgment: The views presented here are those of the authors and do not reflect official policy of Madigan Army Medical Center, the Department of Family Medicine, the Army, or the United States Government.

REFERENCES


Authors’ Response:

Testicular cancer is the most common solid tumor in American men between the ages of 20 and 34, afflicting an estimated three in 1,000.1 Men with a history of cryptorchidism or atrophic testes are at increased risk. The American Cancer Society (ACS) recommends testicular exams for all men as part of a routine cancer-related checkup,2 and the Bright Futures guideline recommends examination of the genitals for problems including testicular cancer for all adolescent males.3 The Institute for Clinical Systems Improvement recommends testicular exams for those at high risk only.4 The American Medical Association supports education in testicular self-examination (TSE) methods for early detection of testicular cancer,3 and the ACS recommends TSE only for those at high risk.2 The US Preventive Services Task Force recommends against screening using either clinical examination or self-examination.6

We are in agreement with much of the commentary in the letter submitted by Drs Seehusen and Edwards. For us, this curriculum serves two main purposes. First, it allows medical students to receive excellent clinical examination training on a sensitive topic with a population (adolescent men) that is relatively neglected in medical school training. Second, it allows students and faculty to discuss the basic principles of screening tests.

This module on testicular cancer is part of a larger curriculum on clinical communication for male cancer screening (www.brown.edu/Research/ICHP/mcshome.shtml). In addition to teaching about exam maneuvers and communication techniques, the module on testicular cancer addresses the screening controversy:

• Do the costs associated with implementing a mass screening effort for testicular care outweigh the benefits? The essential question remains, “Does the procedure produce more harm than good?”

• Critics of testicular cancer screening programs mention lack of sufficient evidence of benefit and the anxiety produced in otherwise healthy young individuals as reasons for not screening.

• Advocates of testicular cancer screening generally present the following arguments in its favor: (1) It’s an inexpensive and simple procedure to perform and teach, (2) The population affected by testicular cancer is very young; any effort to detect cancer that would result in the loss of young lives or years of productivity should be undertaken, (3) Teaching the TSE is a means of introducing topics affecting young men, such as sexually transmitted diseases (STDs) and sexual practices that are difficult to discuss.

We maintain that knowing how to perform and teach a competent testicular exam is an essential medical school competency. The concept of conflicting recommendations is an important one for students beginning their clinical medicine training to understand. The TSE is an excellent example of a screening test with different and changing recommendations from various agencies, and we present it as such.

Julie Taylor, MD, MSc
Catherine Dubé, EdD
Department of Family Medicine
Memorial Hospital of Rhode Island
Pawtucket, RI

REFERENCES


New Research

Assessment of Selected Psychometric Properties of Nonstandard ABFP TIE Scales

To the Editor:
The American Board of Family Practice4 (ABFP) and others2,3 have published various psychometric properties of the widely used ABFP
In-training Examination (ITE) standard scales, eg, composite, pediatrics, surgery. These properties allow the user to make informed decisions regarding the proper use and interpretation of the ABFP ITE standard scale scores. However, Web-based applications have been developed by non-ABFP organizations that allow residency programs to evaluate new nonstandard scales within detailed curricular areas. For instance, one Web site categorizes each ITE question and reports the number of questions a resident missed in each content area. The Web site suggests that the results may be used to modify existing curricula. Approximately 22% of family medicine Accreditation Council for Graduate Medical Education (ACGME)-accredited residencies are registered users of this site.

For the 2003 ITE examination, the Web site listed 191 content areas that were represented by only one ITE test question, eg, irritable bowel syndrome. Twenty-two other content areas were represented by two to 17 questions, eg, there were 17 questions regarding HIV. Thus, content areas were assessed by "scales" containing from one to 17 ITE questions. Whereas the psychometric properties of the standard ABFP ITE scales have been reported, the psychometric properties of these new nonstandard scales are unknown. This study assessed the reliability and positive predictive value (PPV) of these newly offered nonstandard ABFP ITE scales with small numbers of items such as those reported by this Web site.

Methods

We know from psychometric theory that test reliability is greatly influenced by the number of items on the test. In fact, the Spearman-Brown prophecy formula is often used to estimate the change in test reliability as the number of test items increases or decreases. To estimate the reliability of scales reported by the aforementioned Web site, we used the known reliability coefficient (0.87)-1 of the 340-item composite score of the ABFP ITE and the Spearman-Brown prophecy formula. Thus, we estimated the change in the reliability coefficient as the number of items on the ITE decreases from the 340 to 17 or less given that the reliability of the 340-item composite score is 0.87. We then used these reliability estimates and Monte Carlo methodology3 to estimate the PPV.

Results

Decreasing the ITE to 17 items results in an estimated reliability coefficient of 0.25. Decreasing the ITE to one item results in an estimated reliability coefficient of 0.02 for each of the 191 one-item scales. Results indicate that for a 17-item scale with a reliability coefficient of 0.25, the PPV is approximately 27%; for the 172 one-item scales, the PPV is approximately 11%.

Conclusions

If we make inferences about a resident’s fund of knowledge based on an incorrect response to a single item, about 89% of our inferences will be false. Even in the best-case scenario of the 17-item scale, 73% of residents identified as having knowledge deficit in HIV will be false positives. Although it is reasonable to explain the correct and incorrect answers to individual test items, it is entirely unreasonable to make inferences about large domains of knowledge based on scales with small numbers of items, low reliability, and low PPV. Educators should not try to retrofit the standard ABFP ITE and use the results in a way that was not intended. Making changes in curriculum or lecture content based on responses to a single item or small “multi-item” scales is not appropriate.

William H. Replogle, PhD
Department of Family Medicine
William D. Johnson, PhD
Department of Preventive Medicine
University of Mississippi Medical Center

References

2. Replogle WH. Interpretation of the American Board of Family Practice In-training Examination. Fam Med 2001;33(2):98-103.

Medical Decision-making Risk Management Based on Aeronautical Model

To the Editor:

In the field of aviation, it was decided a number of years ago that something different was needed to reduce the number of accidents. Through various studies, it was determined that 85% of all accidents and 52% of fatal general aviation accidents were due to pilot error or poor decision-making risk management. Therefore, the Federal Aviation Administration (FAA) embarked on a process of determining and evaluating the attitudes, behavioral traps, stresses, and other items that enter into decision making by pilots. After 12 years of research, development, and testing, the result was publication in 1987 of manuals oriented to the decision-making needs of pilots. The effectiveness of the process was validated in six independent studies. After pilots received the training, the errors in decision making decreased—leading to a 10% to 50% drop in accidents.

Pilots and physicians have a similar or the same type of makeup, ie, goal oriented, self confident, sense of invulnerability, and macho. Being a physician and a pilot, I thought maybe the malpractice problem could be attacked in the same fashion.

The scope of the malpractice problem in the United States is impressive. According to statistics that