International Journal of Impotence Research (1999) 11, 319–326
 0955-9930/99 $15.00
http://www.stockton-press.co.uk/ijir

Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction

RC Rosen[1]*, JC Cappelleri[2], MD Smith[3], J Lipsky[4] and BM Peña[5]

[1]*University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, Piscataway, NJ;* [2]*Biometrics, Pfizer Central Research, Groton, CT;* [3]*Biometrics, Pfizer Central Research, Sandwich, Kent, UK;* [4]*Marketing Research, Pfizer Pharmaceuticals Group, New York, NY;* [5]*Outcomes Research, Pfizer Pharmaceuticals Group, New York, NY, USA*

An abridged five-item version of the 15-item International Index of Erectile Function (IIEF) was developed (IIEF-5) to diagnose the presence and severity of erectile dysfunction (ED). The five items selected were based on ability to identify the presence or absence of ED and on adherence to the National Institute of Health's definition of ED. These items focused on erectile function and intercourse satisfaction. For 1152 men (1036 with ED, 116 controls) analyzed, a receiver operating characteristic curve indicated that the IIEF-5 is an excellent diagnostic test. Based on equal misclassification rates of ED and no ED, a cutoff score of 21 (range of scores, 5–25) discriminated best (sensitivity = 0.98, specificity = 0.88). ED was classified into five severity levels, ranging from none (22–25) through severe (5–7). Substantial agreement existed between the predicted and 'true' ED classes (weighted kappa = 0.82). These data suggest that the IIEF-5 possesses favorable properties for detecting the presence and severity of ED.

Keywords: erectile dysfunction; diagnostic tests; questionnaires; impotence; erectile function; severity

Introduction

Erectile dysfunction (ED) affects millions of men. The prevalence of ED depends on the population studied and the definition and methods used.[1] Because of this, estimated prevalence rates of ED have varied considerably and have included, for example, 52% from a study in the United States,[2] 32% from a study in the United Kingdom,[3] 26% from a study in Japan,[4] and 19% from a study in Denmark.[5] Despite considerable variation in the populations studied and the methods used, these reports indicate that ED bears a significant correlation with age and a lower quality of life.[6–8]

The International Index of Erectile Function (IIEF), which consists of 15 items and 5 domains, is a psychometrically valid and reliable instrument that was developed through consultations with an international panel of experts for use in determining efficacy of treatment in controlled clinical trials.[9] The IIEF has high sensitivity and specificity for detecting real treatment effects or the lack of treatment effects in patients with ED of broad-spectrum etiology.[9]

There is a need for a simple patient-administered diagnostic tool of ED for easy use by physicians in clinical settings.[10] The original IIEF instrument was designed specifically for use in clinical trials and is not well suited for use as a simple office screening measure. One way to address this need is to investigate an abbreviated version of the IIEF. Erectile dysfunction is a self-reported condition, and there are no objective diagnostic tests available to physicians for confirmation of the condition, making it difficult for physicians to make an accurate diagnosis. Consequently, a need exists for an easy-to-use clinical instrument for the detection of ED that can supplement physical examination and patient history in clinical settings and that can increase the likelihood of a correct diagnosis for men with or without ED.

In the design of an abbreviated questionnaire, a two-thirds reduction in the number of IIEF items, from 15 to 5, was chosen in advance, emanating from discussions with experts and focus groups. The National Institutes of Health (NIH) Consensus Panel has defined erectile dysfunction (ED) as the inability to achieve and/or maintain penile erection sufficient for satisfactory sexual performance.[10] Although

*Correspondence: Raymond C. Rosen, PhD, Department of Psychiatry, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, NJ 08854-5635, USA.
Received 15 February 1999; accepted 21 June 1999

several international definitions of ED exist, we anchored the five items for inclusion based on the NIH definition as it was derived from consensus of a large group of experts. Furthermore, such a diagnostic instrument for ED should discriminate well between men with and without ED and accurately reflect grades of ED severity.[10] This report describes the construction and evaluation of an abbreviated version of the IIEF, designated as the IIEF-5, as a diagnostic tool for ED and provides data supporting its use as a diagnostically valid instrument in clinical settings worldwide.

Methods

Subjects and studies

Men with clinically diagnosed ED were enrolled in four double-blind, placebo-controlled Phase III multicenter clinical trials investigating the safety and efficacy of VIAGRA® (sildenafil citrate). The inclusion criteria for these trials required that patients with ED be at least 18 years old, be in a stable relationship with a female partner for at least the past six months, and have a clinical diagnosis of ED for a minimum of six months. Documented evidence of ED included a medical history of at least six months duration, physician records, and objective testing where available. Patients with ED of broad-spectrum etiology (organic, psychogenic, or mixed organic–psychogenic) were included. Patients with ED were excluded if they had penile anatomical disorders impairing erectile function, major medical illnesses (e.g uncontrolled diabetes or severe renal, hepatic or cardiovascular diseases), major psychiatric disorders, or a history of alcohol or drug abuse. A control group of men without a history of ED was recruited from volunteers at an outpatient community health center. The responses in the studies were from patients in the United States and the United Kingdom.

Selecting the five items

The initial analyses performed for selecting the five items of the IIEF-5 included data from all men in the Phase III trials and all control men who completed the 15 items of the IIEF questionnaire and who had sexual activity or attempted sexual intercourse within the four-week period before completing the questionnaire. This resulted in a total of 1047 men (932 with ED, 115 controls). Analyses of data from patients with ED were restricted to baseline data obtained during enrollment into the clinical trials (that is, before randomization to study drug). Each IIEF item is scored on a five-point ordinal scale where lower values represent poorer sexual function. Thus, a response of 1 for a question was considered the least functional, whereas a response of 5 was considered the most functional.

The statistical program Classification and Regression Trees (CART)[11,12] was used to arrive at a diagnostically optimal set of five items by simultaneously calculating and ranking the relative importance of the 15 items of the IIEF from a multivariate framework in terms of their ability to explicitly discriminate between men with and without ED. The score of each item in the chosen classification tree was summed and scaled relative to the item that best discriminated between ED and no ED. The most important discriminating item received a score of 100 and the least important a score of 0. The five items selected for the IIEF-5 had to clearly discriminate between study subjects with and without ED, as well as address the NIH definition of ED. Possible scores for the IIEF-5 range from 5 to 25.

Evaluating the five items for discrimination

The diagnostic evaluation of the five items chosen for the IIEF-5 included scores from all 1047 subjects who had completed all 15 items of the IIEF. It also included scores from a second cohort of men who had not completed all 15 items but who had provided responses to the five items selected for inclusion in the IIEF-5. Four complementary types of statistical analyses were applied to evaluate the ability of the IIEF-5 to predict whether or not a subject had ED.

First, a receiver operating characteristic (ROC) curve was created in which the true-positive rate and the false-positive rate were paired across all potential cutoff points that distinguished between men with and without ED.[13] In the ROC curve, the true-positive rate (sensitivity) was the proportion of men with ED that the IIEF-5 correctly predicted as having ED; the higher the rate, the better the test. The false-positive rate (1 − specificity) was the proportion of men without ED that the IIEF-5 incorrectly predicted as having ED; the lower the rate, the better the test. The true-negative rate (specificity) was therefore the proportion of men without ED that the IIEF-5 correctly classified as having no ED. Therefore, a desired cutoff score should have a high true-positive rate and a low false-positive rate in partitioning men with ED from men without ED. The ROC curve was determined by comparing a subject's known diagnosis (ED, no ED) with the prediction based on each of the 21 candidate cutoff scores on the IIEF-5 (that is, scores of 5–25). Subjects were retrospectively classified as

having ED if their scores were less than or equal to the candidate cutoff value.

Second, a logistic regression model was analyzed with the Statistical Analysis System (SAS)[14] and was used to further assess how well IIEF-5 scores could predict and relate to the presence or absence of clinical diagnosed ED.

Third, in determining the optimal cutoff score, the CART algorithm[11,12] was used to evaluate the binary split of ED *vs* no ED for each of the 21 possible cutoff points (5–25) on the IIEF-5. Each cutoff point was evaluated on the basis of a goodness-of-split criterion as measured by the Gini diversity (impurity) index.[11,12] The optimal cutoff resulted in the lowest Gini value and the least variability in the actual status of men categorized by CART as having ED or not having ED. The optimal cutoff was based on the assumption, consistent with the literature,[6] that a reasonable estimate of the prevalence rate of minimal-to-complete ED is about 50% in a clinical population of adult men. In CART, this is equivalent to having the error of misclassifying men with ED be equivalent to, or as serious as, the error of misclassifying men without ED.

Fourth, a cross-validation analysis with CART was undertaken to determine the stability of the binary split (ED *vs* no ED) for the optimal cutoff score by developing calibration rules for subsamples containing 90% of the subjects and testing these calibration rules on the other 10% of subjects. The CART algorithm analyzed a total of ten separate subsamples, each with a different 10% of the data used as a test sample, and inspected the classifications for consistency. Cross-validated results were generated as the best estimates that would occur if the selected classification tree was applied to new, fresh data. These cross-validation results were compared with the sample results from the tree applied to all of the sample data to determine the robustness of the evaluation.

For the cross-validated results and the sample results, the sensitivity and specificity were then used to calculate the predicted value positive and the predicted value negative of the IIEF-5.[13] The predicted value positive was the proportion of men with actual ED among men classified as having ED based on their IIEF-5 scores. The predicted value negative was the proportion of men without ED among men classified as not having ED based on their IIEF-5 scores. These predicted values depended on the prevalence rate of ED in a clinical population, assumed to be about 50%.[6]

Exact 95% confidence intervals for sensitivity, specificity, predicted value positive, and predicted value negative were calculated using StatXact.[15] For the cross-validated and sample results, the kappa coefficient and its 95% confidence intervals were also calculated[15] to provide a measure of overall agreement, beyond chance, between the actual diagnosis and the diagnosis predicted from the IIEF-5 scores.

Evaluating the five items for classifying degrees of erectile dysfunction

Men were not asked to self-rate their severity of ED, nor did clinicians rate the overall severity of ED of their patients. Because there was no independent criterion for severity of ED, we created a surrogate measure for it by choosing the IIEF item not in the IIEF-5 that best discriminated ED from no ED and that conformed to the NIH definition of ED. After determining this surrogate item, we established a gradient of severity for ED by adopting the following 3-tier approach:

Step 1. The clinical outcome (ED, no ED) was expressed as a function of the five ordinal values on the surrogate item for ED severity to determine, using CART, which categories of this item belong to men with ED and which belong to men without ED.

Step 2. For men with actual ED, categories of the item that belonged to ED (from *Step 1*), which were analyzed as ordinal responses, were expressed as a function of IIEF-5 scores to determine, using CART, a range of IIEF-5 scores that corresponded to each category of the item. The range of IIEF-5 scores for no ED was determined beforehand by those scores exceeding the optimal cutoff point from the analysis that discriminated between ED and no ED (see previous section).

Step 3. Using a general linear model and Scheffe's multiple comparisons test in SAS,[14] we tested whether the mean IIEF-5 score of a given severity category of ED significantly differed from each of the other categories.

The weighted kappa coefficient[15] was calculated to provide an indication of overall agreement, beyond chance, between the 'true' classification of ED severity (from categories on the surrogate item for ED severity) and the predicted classification of ED severity (from the IIEF-5 scores). In addition, the correlation between these two sets of classifications were computed using the Spearman rank-order correlation and the Kendall rank-order (tau-b) correlation; exact 95% confidence intervals of these correlations were calculated.[15]

Results

Selecting the five items

Selecting the five items for inclusion in the IIEF-5 was accomplished by reviewing the 1047 responses of 932 men with ED and 115 men without ED who

Table 1 Relative importance of the items on the IIEF to discriminate between men with and without ED

IIEF[a] *Item*		*Relative importance (100 = most important, 0 = least important)*
5	Maintenance ability	100.0
15	Erection confidence	95.4
4	Maintenance frequency	88.1
2	Erection firmness	85.3
3	Penetration ability	79.6
7	Intercourse satisfaction	67.4
1	Erection frequency	9.2
6	Intercourse frequency	3.6
11	Desire frequency	3.5
10	Orgasm frequency	2.2
13	Overall satisfaction	0.9
14	Relationship satisfaction	0.6
12	Desire level	0.4
9	Ejaculation frequency	0.3
8	Intercourse enjoyment	0.0

[a] IIEF = International Index of Erectile Function.

completed the full-length, 15-item IIEF questionnaire. For these men, the average age in the ED group (54.1 y) and the control group (56.1 y) was similar. For subjects with ED, the average duration was 4.2 y and their etiologies were 49.1% organic, 22.0% psychogenic, 28.6% mixed, and 0.3% other. Table 1 shows the rank in relative importance of items on the 15-item IIEF questionnaire in discriminating between men with and without ED.

As shown in Table 1, the first four discriminators—maintenance ability (Item 5), erection confidence (Item 15), maintenance frequency (Item 4), and erection firmness (Item 2)—conform to the NIH definition of ED. The 'satisfactory sexual performance' component of the NIH definition was not explicitly characterized by the top five items with the greatest discriminatory power. Therefore, Item 3, the fifth best discriminator, was replaced with Item 7 on sexual intercourse satisfaction, the next best discriminator. Hence, the IIEF-5 consists of Items 5, 15, 4, 2, and 7 from the IIEF. Table 2 presents the components of the IIEF-5 questionnaire. Items have been rephrased to address the past six months of sexual activity, instead of the past four weeks, to conform to the NIH's current reference period for establishing a diagnosis of ED.

Evaluating the five items

After the items on the IIEF-5 were chosen, an evaluation of the questionnaire was performed based on all men with completed scores on the IIEF-5: 1036 men had ED and 116 men did not have ED. These 1152 subjects came from the same set of clinical studies from which the initial 1047 subjects were selected and included an additional 105 subjects from these studies who had provided responses to IIEF-5 items but did not respond to at least one of the nine remaining items on the 15-item IIEF questionnaire. Ten men with ED who reported no sexual activity during the designated period were not included in the analysis. The average IIEF-5 score for men clinically diagnosed with ED (mean ± s.d.: 11.1 ± 4.7, range = 5–25) was significantly lower ($P < 0.0001$) than the average IIEF-5

Table 2 The IIEF-5 questionnaire[a]

Over the past six months:						
1	How do you rate your **confidence** that you could get and keep an erection?	Very low	Low	Moderate	High	Very high
		1	2	3	4	5
2	When you had erections with sexual stimulation, how often were your erections hard enough for penetration?	Almost never/never	A few times (much less than half the time)	Sometimes (about half the time)	Most times (much more than half the time)	Almost always/always
		1	2	3	4	5
3	During sexual intercourse, **how often** were you able to maintain your erection after you had penetrated (entered) your partner?	Almost never/never	A few times (much less than half the time)	Sometimes (about half the time)	Most time (much more than half the time)	Almost always/always
		1	2	3	4	5
4	During sexual intercourse, **how difficult** was it to maintain your erection to completion of intercourse?	Extremely difficult	Very difficult	Difficult	Slightly difficult	Not difficult
		1	2	3	4	5
5	When you attempted sexual intercourse, how often was it satisfactory for you?	Almost never/never	A few times (much less than half the time)	Sometimes (about half the time)	Most times (much more than half the time)	Almost always/always
		1	2	3	4	5

[a] The IIEF-5 score is the sum of the ordinal responses to the five items; thus, the score can range from 5 to 25.

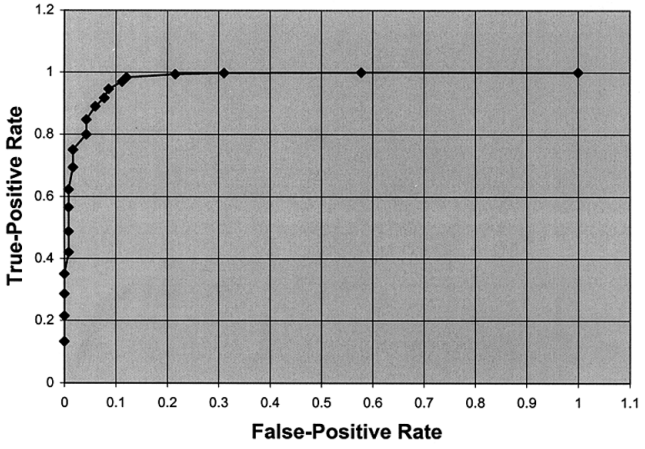

Figure 1 Receiver operating characteristic curve for the IIEF-5.

score for men without ED in the control group (23.3 ± 2.8, range = 9–25). The average age was similar between the two groups (55.8 y for men with ED, 54.1 y for controls).

Figure 1 depicts the ROC curve for the IIEF-5. This curve, which lies predominantly far upward and to the left, shows relatively high true-positive rates and low false-positive rates. The estimated area under the ROC curve was 0.97. Hence, from a randomly selected pair of men, one with ED and one without ED, there was about a 97% chance of correctly identifying the man with ED and the man without ED based on their IIEF-5 scores alone.

The ROC curve can help determine desired cutoff points on the IIEF-5 for predicting a diagnosis of ED. As shown in Table 3, integer scores ranging from 14 to 22 on the IIEF-5 were candidate cutoff points (based on their sensitivity and specificity being at least 0.75) that provided adequate discrimination for distinguishing between men with ED and those without ED. Furthermore, a logistic regression model indicated that for every 1-point increase in a subject's IIEF-5 score, the odds of having ED (relative to not having ED) decreased by about half (odds ratio = 0.48; 95% confidence interval (CI), 0.42–0.54).

The CART analysis indicated that a score of 21 was the optimal cutoff score for determining whether or not a subject had ED. Men who scored ≤ 21 were retrospectively classified as having had ED, whereas men who scored > 21 were retrospectively classified as not having had ED. Table 4 contains the diagnostic statistics for the sample data and the cross-validation, which were identical in this case, for an optimal cutoff point of 21.

The corresponding sensitivity was 0.98 (that is, 1018 men predicted to have ED out of 1036 men clinically diagnosed with ED) and the specificity was 0.88 (that is, 102 men predicted not to have ED out of 116 controls). Thus, for men who had ED, 98% of them were classified correctly by the IIEF-5 as having had ED; for men who did not have ED, 88% of them were classified correctly by the IIEF-5 as not having had ED. These rates, coupled with an estimated clinical rate of ED of 50%, suggest that there was an 89% chance that a subject actually had ED given that he was diagnosed as much based on his IIEF-5 score (that is, 0.89 is the positive predictive value). There was about a 98% chance that a subject did not actually have ED given that he was diagnosed as not having had ED based on his IIEF-5 score (that is, 0.98 is the negative predictive value). The estimated kappa coefficient of 0.85 indicated substantial agreement, above and beyond

Table 4 Diagnostic statistics for the IIEF-5 cutoff point of 21 for the sample data and for the cross-validation

	Estimate	*95% Confidence Interval*
Sensitivity	0.98	0.97–0.99
Specificity	0.88	0.80–0.93
Predicted Value Positive	0.89	0.86–0.91
Predicted Value Negative	0.98	0.96–0.99
Kappa	0.85	0.80–0.90

Table 3 True- and false-positive rates and predictive probabilities for candidate cutoff scores[a]

Cutoff Score	*True-positive rate (Sensitivity)*[b]	*False-positive rate (1 − Specificity)*[c]	*Predictive probability of ED for a score ≤ cutoff score*[d]
14	0.75	0.02	1.00
15	0.80	0.04	0.99
16	0.85	0.04	0.98
17	0.89	0.06	0.97
18	0.92	0.08	0.94
19	0.94	0.09	0.88
20	0.97	0.11	0.78
21	0.98	0.12	0.63
22	0.99	0.21	0.45

[a] Men with a IIEF-5 score less than or equal to a given cutoff score were retrospectively classified as having had ED, whereas men with scores greater than the same cutoff were retrospectively classified as not having had ED. IIEF-5 scores can range from 5 to 25. Probabilities are rounded.
[b] The true-positive rate, or sensitivity, is the proportion of men with ED who were classified correctly as having had ED.
[c] The false-positive rate, or 1 minus the specificity, is the proportion of men without ED who were classified incorrectly as having had ED.
[d] This is the (unconditional) probability of ED for men with scores less than a given cutoff score, where lower scores suggest more ED.

Table 5 Number of men classified by 'true' grading and predictive grading for erectile dysfunction

	Predictive grading of ED					
'True' grading of ED[a]	*Severe*	*Moderate*	*Mild to moderate*	*Mild*	*No ED*	*Total*
Severe	224	54	10	0	0	288
Moderate	50	134	57	5	0	246
Mild to moderate	17	56	98	15	0	186
Mild	3	32	63	44	3	145
No ED	0	0	0	3	100	103
Total	294	276	228	67	103	968

[a] As specified in the text, a 'true' grading was based on a category of penetration ability (item 3) of the IIEF, whereas a predictive grading was based on a given range of IIEF-5 scores as specified in the text. ED = erectile dysfunction.

chance, between clinical diagnosis and predicted diagnosis.

Classifying severity of erectile dysfunction

Step 1. Based on the criteria outlined in the Methods section, Item 3 of the IIEF—frequency of penetration for sexual intercourse— was selected as the proxy measure for severity of ED. Men who responded with a 1 (almost never/never), 2 (a few times (much less than half the time)), 3 (sometimes (about half the time)), or 4 (most times (much more than half the time)) concerning how often they were able to penetrate or enter their partner were classified as having ED, whereas those who responded with a category code of 5 (almost always/always) were classified as not having ED.

Step 2. Response categories for Item 3 of the IIEF were best related to the IIEF-5 scores as follows: the response 'almost never/never' of Item 3 corresponded to an IIEF-5 score between 5 and 7 (inclusive), 'a few times' to a score between 8 and 11, 'sometimes' to a score between 12 and 16, 'most times' to a score between 17 and 21, and 'almost always/always' to a score between 22 and 25. This last category was predetermined from the previous analysis that indicated that a score above 21 suggested no ED. Thus, ED severity was classified into the following five categories based on IIEF-5 scores; severe (5–7), moderate (8–11), mild to moderate (12–16), mild (17–21), and no ED (22–25). These categories were based on our clinical understanding of the scores.

Step 3. The average IIEF-5 scores for men with ED increased, as expected, with increased penetration frequency: Category 1, 6.5 (standard deviation (s.d.) = 2.0); Category 2, 9.9 (s.d. = 2.8); Category 3, 12.3 (s.d. = 3.2); Category 4, 14.5 (s.d. = 3.6). The mean was 24.1 (s.d. = 1.2) for men without ED who responded to Category 5. After the global null hypothesis of equal IIEF-5 scores across the five categories of ED severity was rejected ($P < 0.001$), each of the ten pairs of five category means was compared and was significantly different according to Scheffe's 95% confidence intervals.

Table 5 provides the number of men classified by 'true' grade and predictive grade of ED severity. A total of 968 men gave complete and non-zero responses to Item 3 on the IIEF and to the five items on the IIEF-5. Men tended to be classified correctly into their 'true' level of ED more often than any other level. An exception was men with a 'true' mild level of ED; they were classified most frequently into the predictive grade of 'mild to moderate.' Men who were misclassified tended to be assigned into a degree category immediately adjacent to the 'correct' category, rather than into a more remote category. Substantial overall agreement between 'true' and predictive grades was found; the weighted kappa value was 0.82 (95% CI, 0.80–0.85). The Spearman and Kendall (tau-b) rank-order correlations between these grades, respectively, were 0.79 (95% CI, 0.76–0.82) and 0.72 (95% CI, 0.69–0.75), suggesting a high association.

Discussion

The 15 items of the IIEF can be segregated into five sexual function domains: erectile function, orgasmic function, sexual desire, intercourse satisfaction, and overall satisfaction.[9] It is not surprising that four of the five items on the IIEF-5 were taken from the six-item erectile function domain, which has the strongest psychometric properties of the five domains the IIEF[9] and can itself serve as a complementary diagnostic measure.[16,17] But this finding could not have been ascertained unless the current investigation was undertaken. Of the five items on the IIEF-5, only Item 7 (intercourse satisfaction) is not a member of the erectile function domain. Hence, the results of this report also lend support for items in the erectile function domain as an additional diagnostic tool. Moreover, the diagnostic results between the erectile function domain and the IIEF-5 are nearly identical (data not shown).

There are a few methodologic differences, however, between these two diagnostic tools. The inclusion of Item 7 on sexual intercourse satisfaction is a distinguishing feature of the IIEF-5 that is absent from the erectile function domain. Intercourse satisfaction gives a high level of discrimination and addresses a central element in the NIH definition of ED.[10] Item 7 was used to establish a gradient of ED severity for the erectile function domain, whereas penetration ability (Item 3) served this key role for the IIEF-5. The IIEF-5 has only five items, whereas the erectile function domain has six items; this may or may not increase the number of completed responses in clinical settings but is unlikely to reduce it. Finally, items in the IIEF-5 are phrased to reference the prior six-month period, which conforms with the NIH's current reference period for establishing a diagnosis of ED. In contrast, items in the erectile function domain have a reference period of four weeks.

The ROC curve provides evidence that the IIEF-5 can be an excellent diagnostic test. Although several cutoff scores may be acceptable (see Table 3), the cutoff score of 21 is optimal because it possesses the best discriminatory diagnostic properties and is based on a reasonable assumption that, in a clinical setting, misclassifying a subject with ED is as serious as misclassifying a subject without ED. This assumption, which is an important consideration in a clinical population, neutralizes the classification process and removes the subjectivity of quantifying and favoring one type of misclassification over the other, which may vary from clinician to clinician. Nonetheless, further research would be useful to extend the primary diagnostic analysis by determining optimal cutoff scores for different prevalence rates of ED and unequal misclassification rates. Research along these lines may expand the utility of the IIEF-5 by increasing the precision of diagnostic cutoffs for clinical subpopulations with varying levels of risk for ED and developing its use as an epidemiologic instrument to estimate prevalence rates of ED in these subgroups.

The evaluation of the IIEF-5 was retrospective because the clinical diagnosis of whether or not a subject had ED was determined before, rather than after, consideration of his score. The evaluation was based on five items that were part of a larger pool of 15 items that were all completed together; responses to the five items were not evaluated in isolation but in tandem with the other ten items on the IIEF. Moreover, the current NIH guidelines recommend using six months, instead of the four weeks used in this investigation, as the reference period for answering questions on ED. The best way to ascertain the impact of these modifications would be to incorporate them into a prospective study for validation. A similar type of validation approach was adopted when a diagnostic aid was first evaluated retrospectively for predicting myocardial infarction in patients arriving at the emergency room with acute chest pain.[18] Further prospective research led to validation of the same diagnostic instrument.[19] Thus, as a further evaluation of the utility of the IIEF-5 as a diagnostic tool, a number of prospective US and European studies of ED will adopt this measure.

Additional research is needed on the validity of our proposed severity classification for ED. This research should include a validation study that assesses the degree of agreement and magnitude of correlation between subject self-assessment of erectile function and the IIEF-5 with respect to levels of ED severity. Further research is also needed on the relationship between ED severity and subjective distress or quality of life effects in different patient groups.[6] The current study was limited to the assessment of erectile function and intercourse satisfaction. Further studies are underway on the relationship of these variables to patient quality of life.

The diagnostic evaluation of the IIEF-5 was based only on men who reported having attempted sexual intercourse in the four-week period before filling out the questionnaire (that is, those men with non-zero coded responses). When men who reported not having attempted sexual intercourse are included in the classification of severity, resulting in a total of 979 subjects, the same level of overall agreement between predicted and 'true' levels of ED was obtained (weighted kappa = 0.82). If men who reported having had no sexual activity are considered, we suggest that the severest category of ED be graded from 1 to 7, instead of 5 to 7, provided that they had clinically diagnosed ED or were involved in a stable relationship with a female partner.

With the advent of recently available oral therapeutics for the treatment of ED,[20,21] the need for accurate diagnoses within worldwide subpopulations of men is greater than ever. The IIEF-5 is intended to complement, not supplant, clinical judgment and useful diagnostic assessments. It may be particularly useful as an initial screening instrument in a general practice setting. One company (Pfizer), for instance, has used it as a Sexual Health Inventory for Men. Because of its simplicity and the favorable diagnostic properties reported here, the IIEF-5 could aid in decreasing the incorrect diagnosis of ED and decreasing the number of undiagnosed cases of ED worldwide.

Acknowledgments

These studies were funded by Pfizer Inc. The authors, four as employees (JCC, MDS, JL, and BMP) and one as a consultant (RCR), have a financial arrangement with the company whose product is mentioned in the article. Although data

are used from Viagra® clinical trials, all analyses, results, and conclusions are based on pre-treatment (baseline) data and did not involve any treatment intervention.

The authors gratefully acknowledge Dr. Peter A. Rittenhouse and Dr. Michael Sweeney for aid in the preparation of this manuscript.

References

1 Melman A, Gingell JC. The epidemiology and pathophysiology of erectile dysfunction. *J Urol* 1999; **161**: 5–11.
2 Feldman HA *et al*. Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study. *J Urol* 1994; **151**: 54–61.
3 Spector KR, Boyle M. The prevalence and perceived aetiology of male sexual problems in a non-clinical sample. *Br J Med Psychol* 1986; **59**: 351–358.
4 Shirai M *et al*. A stochastic survey of impotence population in Japan. *Impotence* 1987; **2**: 67–93.
5 Solstad K, Hertoft P. Frequency of sexual problems and sexual dysfunction in middle-aged Danish men. *Arch Sex Behav* 1993; **22**: 51–58.
6 Benet AE, Melman A. The epidemiology of erectile dysfunction. *Urol Clin North Am* 1995; **22**: 699–709.
7 Jønler M *et al*. The effect of age, ethnicity and geographical location on impotence and quality of life. *Br J Urol* 1995; **75**: 651–655.
8 Laumann E, Paik A, Rosen RC. Sexual dysfunction in the United States: prevalence and predictors. *JAMA* 1999; **281**: 537–544.
9 Rosen RC *et al*. The International Index of Erectile Function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology* 1997; **49**: 822–830.
10 NIH Consensus Development Panel on Impotence. *JAMA* 1993; **270**: 83–90.
11 Breiman L, Friedman J, Olshen R, Stone C. *Classification and Regression Trees. Wadsworth: Pacific Grove,* 1984.
12 Steinberg D, Colla P. *CART. Tree-Structured Non-Parametric Data Analysis.* Salford Systems: San Diego, 1995.
13 Weinstein MC *et al*. *Clinical Decision Analysis.* WB Saunders Company: Philadelphia, 1980, pp 75–130.
14 SAS Institute Inc. *SAS/STAT® User's Guide,* Version 6, Vol 2, 4th ed.: SAS Institute: Cary, 1990.
15 *StatXact 3 for Windows Statistical Software for Exact Nonparametric Inference User Manual.* CYTEL Software: Cambridge, MA, 1995, pp 399–401, 639–663, 675–680.
16 Cappelleri JC *et al*. Some developments on the International Index of Erectile Function (IIEF). *Drug Inf J* 1999; **33**: 179–190.
17 Cappelleri JC *et al*. A diagnostic evaluation of the Erectile Function Domain of the International Index of Erectile Function (IIEF). *Urology* 1999; **54**: 346–351.
18 Goldman L *et al*. A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *N Engl J Med* 1982; **307**: 588–596.
19 Goldman L *et al*. A computer protocol to predict myocardial infarction in emergency department patients with chest pain. *N Engl J Med* 1988; **318**: 797–803.
20 Goldstein I *et al*. Oral sildenafil in the treatment of erectile dysfunction. *N Engl J Med* 1998; **338**: 1397–1404.
21 Eardley I. New oral therapies for the treatment of erectile dysfunction. *Br J Urol* 1998; **81**: 122–127.