Screening for Ophthalmic Disease in Older Subjects Using Visual Acuity and Contrast Sensitivity

Russell L. Woods, PhD,1 Stephen J. Tregear, DPhil,1 Reginald A. Mitchell, PhD2

Objective: Despite early interest in contrast sensitivity as a screening test for ophthalmic disease, most published opinion suggests that there is no benefit over conventional measurement of visual acuity. Taking a primary care perspective of screening, the authors evaluated the ability to discriminate those with any diagnosed ophthalmic disease in a large sample representative of the general population.

Design: Retrospective analysis of a clinical, cross-sectional survey. Snellen visual acuity, contrast sensitivity (Arden plates, American Optical contrast sensitivity test), and ophthalmic diagnosis were reported previously.

Participants: A sample of 3283 subjects, all aged at least 50 years, were selected randomly from residents of a health district in Sydney, Australia. Ophthalmologic diagnosis (ophthalmic disease presence/absence) had been confirmed for 2522 of these subjects.

Main Outcome Measures: Signal detection techniques (the receiver-operating characteristics function [ROC], quality ROC [QROC], and weighted kappa coefficient of association [κ]) were used to evaluate test discriminability.

Results: Because analyses of right and left eyes were almost identical, only right eye results are presented. Advantages of κ over ROC were shown. Discrimination of those with diagnosed ophthalmic disease from those without ophthalmic disease was best with Arden plate 7 (κ0.5 = 0.93) and was better than distance Snellen visual acuity (κ0.5 = 0.59). Arden plate 7 (6.4 cyc/deg) correctly assigned 96% of subjects at its optimal pass–fail criterion.

Conclusions: In the primary care setting, a person older than 50 years of age with reduced contrast sensitivity, as determined by Arden plate 7, requires extra care in subsequent examinations because this person is likely to have an ophthalmic disease. Ophthalmology 1998;105:2318–2326

Escalating costs of healthcare provision are causing concern in many countries. Screening for disease is one method used to reduce costs. Effective screening ensures appropriate and timely healthcare for those in need while ensuring that those who do not require a particular service do not congest the healthcare system. Screening can occur at many levels. For example, primary care screening may be identification of those worthy of further examination. At higher levels, the distinction between screening and diagnosis may not be clear. One potential screening test that has been evaluated in a number of studies is contrast sensitivity.1–11 However, whether such tests have a role within clinical practice remains unclear. Consequently, contrast sensitivity testing has not been adopted as a common clinical measure,12 despite the production of many different clinical contrast sensitivity tests.13

The clinical use of contrast sensitivity tests has been confounded by the fact that it is affected by many ophthalmic diseases and conditions,1–11,14–18 but not all patients with such conditions show changes in contrast sensitivity.9,10,16 Furthermore, not all patients with the same condition show the same changes in contrast sensitivity.19–22 and patients with different ophthalmic diseases may show apparently similar changes in contrast sensitivity.14,15,17 The poor reliability of many measures of contrast sensitivity suggests that the ability to detect, diagnose, or monitor ophthalmic disease will be restricted.23–30

One of the earliest roles envisaged for contrast sensitivity testing was screening for ophthalmic diseases such as glaucoma.1 Despite these early predictions, no study has demonstrated sufficiently high discriminability to make general screening with any contrast sensitivity test acceptable in clinical practice.15 High correlations between contrast sensitivity and visual acuity limit the potential benefits of contrast sensitivity.25 Consequently, contrast sensitivity is perceived as supplementary to visual acuity.30 Despite these sentiments, there have been few attempts to investigate the ability to screen for ophthalmic disease in a general popu-
lation, and the implicit suggestion that visual acuity is better at discriminating ophthalmic disease than contrast sensitivity has not been demonstrated clearly.

The failure to demonstrate reasonable discriminability between those with ophthalmic pathology and those without, using contrast sensitivity, may be because of limitations in the approach of previous studies rather than limitations in the test itself. These limitations include the method of scoring and using the results of a multidimensional test, confusion over use of multi-eye data, the method (or lack of a method) of determining the optimal screening criterion, the selection of subjects (e.g., referent population), and the choice of condition(s) for which the test screens. We argue that contrast sensitivity has the most potential as a screening test when used in a primary care setting (e.g., general medical practice) to discriminate those with no ophthalmic disease from those with ophthalmic disease. In this primary care perspective, screening indicates those individuals who are likely to have an ophthalmic disease and who are therefore worthy of additional investigation.

To evaluate whether contrast sensitivity and visual acuity had a role in primary care screening for ophthalmic disease, we conducted a retrospective study using information from a previously published study.18,32 To evaluate the optimal criteria for screening, we used the weighted kappa function, derived from the weighted kappa coefficient of association, which we recommend as the appropriate measure of discriminability. Despite our misgivings, our analysis suggests that contrast sensitivity is a far better primary care screening test than has been reported previously.

**Patients and Methods**

Mitchell selected subjects from a random sample of residents at least 50 years of age in the Western Metropolitan Health Region of New South Wales, Australia, over the period from 1988 to 1990. From those older than 64 years of age, a random sample of 2.5% of residents was selected. To retain a manageable sample size, a random sample of 1% of residents 50 to 64 years of age was selected. The projected sample size was 3251 persons from the source population.

Mitchell made initial contact with subjects through their general medical practitioner. Informed consent was obtained. Subjects provided information about their ophthalmic health and details of their ophthalmic practitioner. After appropriate written consent, the ophthalmic practitioner was contacted and the reported ophthalmic conditions were confirmed. Principal diagnosis was defined as the ophthalmic condition that, in the opinion of the ophthalmic practitioner, was the principal cause of visual impairment. The incidence of multiple ophthalmic disease diagnosis was not recorded.

A total of 3283 subjects, comprising 1478 males and 1805 females, were included in the study by Mitchell. The ophthalmic status of 118 individuals could not be confirmed because they had not attended an ophthalmic practitioner in the preceding 3 years. Of the remaining 3165 subjects, the ophthalmic diagnoses by an ophthalmologist could be confirmed for 2522 subjects. The remainder were made by other ophthalmic practitioners (n = 545), or confirmation of principal diagnosis by the ophthalmologist was not clear (n = 98). For the remainder of this report, we consider only those 2522 subjects with confirmed ophthalmologic diagnosis.

Because we wished to evaluate the ability of the tests to screen for ophthalmic disease within the entire population of interest (persons at least 50 years of age), the population sample was biased to older subjects (2.5% vs. 1%), and the prevalence of diagnosed ophthalmic disease was greater in the older subjects (Table 1), the data were corrected for this sampling artifact. On the assumption that the sample of the “younger” subjects was representative of the general population, we multiplied the frequencies of younger subjects (50–64 years) within each category by a factor of 2.5. For example, in the case of subjects with diagnosed ophthalmic disease and a score of 2 units on Arden plate 6, the number expected if sample density had been 2.5% in both groups was 2.5 × 48 + 63 = 183. Interestingly, despite age-related differences in diagnosis prevalence (Table 1), this sample density correction made no appreciable difference to the discriminability analyses.

**Vision Measurement**

Mitchell measured visual acuity and contrast sensitivity monocularly with each eye using the most recent, appropriate optical correction (e.g., spectacles) for the viewing distance (one examiner). Visual acuity was measured at distance (6 m) using a Snellen chart and at near (40 cm) using a reduced Snellen chart with the last fully completed line recorded as the acuity. Visual acuity worse than 6/60 was recorded as CF (count fingers), HM (hand movements), or PL (perceived light). Unfortunately, the smallest line of letters on the charts was 6/6, since Elliott et al reported that the average visual acuity of subjects without ophthalmic disease slowly decreased from approximately 6/5 (−0.10 ± 0.06 logarithm of the minimum angle of resolution [logMAR]) at 50 to 54 years to approximately 6/6 (−0.02 ± 0.05 logMAR) at 75 to 80 years of age. Further, because the second smallest line was 6/9, the ability to detect small, but clinically significant, differences in acuity was limited.

Contrast sensitivity was measured at 57 cm using the Arden plates. The six plates assess contrast sensitivity at spatial fre-
Discriminability and Optimal Criterion Assessment

Vision measurements from right and left eyes were analyzed separately. Signal detection techniques (refer to Appendix) (the receiver-operating-characteristics (ROC), \( r \), the quality ROC (QROC), \( k \), and the weighted kappa function \( r \) derived from the weighted kappa coefficient of association, \( k \),) were used to evaluate the discriminability and to determine the optimal decision criterion (test score), \( t_c \), of the vision measures. Examples of the frequency distribution and associated functions for distance Snellen visual acuity are shown in Figure 1.

In our analysis of the discriminability of contrast sensitivity and visual acuity, we calculated the slope of the ROC function and evaluated the QROC and weighted kappa functions. Because information on the costs of incorrect diagnoses was not available, we used a weighting \( r = 0.5 \) (relative costs of decisions were equal). Normality of frequency distributions was assessed using the Kolmogorov–Smirnov test.\(^{33}\)

Results

Ophthalmic Disease

Because results for right and left eyes were very similar, we report only right eye results. On the basis of the confirmed ophthalmologic diagnosis, 48% of eyes were normal (no ophthalmic disease). Ophthalmologic examination identified a wide range of ophthalmic disease diagnoses (Table 1). Because only principal ophthalmologic diagnosis was recorded, Table 1 does not represent the true prevalence of each condition. The prevalence of (any) diagnosed ophthalmic disease increased from 0.39 in the 50- to 64-year age group, to 0.73 in those 65 to 74 years, and 0.86 in those older than 74 years of age.

Vision Scores

Vision scores were not normally distributed. On average, those with age-related maculopathy, cataract, and glaucoma had a sig-

\[ K_{1}, K_{0}, K_{0.5} \]

Figure 1. An example of the method of data analysis using Snellen visual acuity measured at 6 m with the right eye. Results for the left eyes were almost identical. (A) Frequency distributions for the nondiseased and diseased groups adjusted for age-related sample density; (B) receiver-operating characteristics function; (C) quality receiver-operating characteristics function; (D) weighted kappa functions \( K_{1} \), \( K_{0.5} \), and \( K_{0} \). Significantly reduced visual acuity (Wilcoxon–Mann–Whitney test, \( P < 0.0001 \)) and a significantly reduced contrast sensitivity at the four higher spatial frequencies compared with those with no diagnosed ophthalmic disease.\(^{18}\)

Discriminability and Optimal Criteria

Because our goal was to evaluate the ability to screen for ophthalmic disease in a primary ophthalmic care setting (i.e., detection not diagnosis), we investigated the ability of the measured visual acuity and contrast sensitivity to detect any disease condition identified by ophthalmologic diagnosis. Because not all of the reported principal ophthalmic disease diagnoses would be expected to produce an effect on vision (e.g., pterygium, pseudoxefoliation), this was a harsh, but realistic test in a representative population. Inclusion of minor ophthalmic conditions would lower sensitivity. Because there were no appreciable differences in the discriminability analyses between the two eyes (even though diagnosed ophthalmic disease among those 50–64 years of age was significantly more likely in the left eye: 0.39 vs. 0.43, McNemar change test, \( P < 0.0001 \)), only right eye results are reported.
between those with and without disease for a given r in the right eye with (plate 7. The higher the k and the Modified Arden score (Table 2). The best discrimination in each eye was found with the Arden plate 7 and the Modified Arden score (Table 2).

Because there was an increase in the variance of vision perfor-

tion. The best discrimination in each eye was found with the Arden plate.

All of the vision measures were found to discriminate, to varying degrees, those with a diagnosis of ophthalmic disease from those without. The optimal decision criterion found using ROC analysis and QROC analysis was similar to that found with the weighted kappa function, and therefore these results are not reported separately. This occurred because, for this sample, the prevalence was almost 0.5 and in the absence of other information, we have presumed that the cost of decisions was approximately equal for those with and without disease diagnosis (r = 0.5). Examples of the weighted kappa function for visual acuity are shown in Figure 1D and for Arden plates 2, 5, and 7 in Figure 2. The best discrimination was found with Arden plate 7 and the Modified Arden score (Table 2).

Effect of Age on Discriminability

Because there was an increase in the variance of vision performance measures (95% confidence limits) with increasing age and disease prevalences differed with age (Table 1), different tc may be appropriate at different ages. Analysis of our data by age indicated that the tc for all tests was the same for each age group. For all of the contrast sensitivity measures, the discriminability remained virtually unchanged. For visual acuity, the discriminability was lower for the 50- to 64-year-old subjects (e.g., distance visual acuity: $k_{0.5} = 0.51$). The lack of an age-related change in tc does not mean necessarily that the tc found in our study (age $\geq 50$ years) will be optimal for younger groups.

Interocular Difference

Another potential method for evaluating the presence of ophthalmic disease is the absolute difference in vision score between the two eyes (interocular difference). This method of discrimination, which is common in clinical practice and has been suggested as discriminating those with glaucoma from those without, presumes that the effect of ophthalmic disease on visual performance progresses at different rates in the two eyes.

To conduct these analyses, it was necessary to assign ranks to the Snellen visual acuity scores. Hence, for those with visual acuity that could be measured on the Snellen chart (i.e., $\leq 6/60$), the difference was in “lines” of visual acuity. When both eyes could not see the grating on an Arden plate or visual acuity was light perception in both eyes, the difference in vision could not be assigned meaningfully, and these cases were treated as a maximum difference score. This was a reasonable treatment since a patient with such poor vision in both eyes generally would be considered worthy of further investigation.

Interocular difference was a poor discriminator of those with diagnosed ophthalmic disease in one or both eyes. In keeping with clinical wisdom, a visual acuity difference of one rank (generally equivalent to one line of Snellen visual acuity) was the optimal discrimination criterion, but the $k_{0.5}$ was low (e.g., distance visual acuity, $k_{0.5} = 0.41$). An interocular difference in contrast sensitivity on any of the Arden plates was a poor discriminator of diagnosed ophthalmic disease. Again, the best of these was plate 7 ($t_c = 3; k_{0.5} = 0.23$).

Preselection for Reduced Visual Acuity

Reduced visual acuity, without contrast sensitivity measurement, often is considered sufficient evidence to warrant further clinical examination. In addition, it has been suggested that contrast sensitivity be measured only when the patient reports reduced vision and visual acuity is normal.$^{13,15,30,45}$ To examine these approaches in the context of screening, we chose three visual acuity levels for preselection: (1) severely reduced visual acuity$^{36}$ (e.g., visual acuity $< 6/60$); (2) low vision$^{36}$ (visual acuity $< 6/18$); and (3) visual acuity $< 6/6$. Hence, all subjects with distance visual acuity below the preselection criterion were considered to have tested positive and not to require contrast sensitivity measurement, and Arden plate scores were used for all other subjects.

In the first two analyses, the $t_c$ values were not modified by preselection and discriminability was unaltered or slightly reduced. In the third analysis, the preselection forced the discriminability of visual acuity alone to form a lower limit. Plates 2, 3, and 4 did not improve the discriminability over that achieved with visual acuity alone. The discriminability of the other three plates was better than the visual acuity criterion alone but was not as good as each Arden plate on its own.
Combining Vision Measures to Predict Ophthalmic Disease

Bayesian theory suggests that given two unrelated measures, which can each discriminate disease, discriminability is increased by using both tests. This effect of Bayesian statistics is used regularly by diagnosticians, a positive test result with both tests indicating a greater likelihood of the presence of disease. We examined both the “OR” (fail test 1 or test 2) and “AND” (fail both test 1 and test 2) conditions. Arden plates were combined with visual acuity or another Arden plate.

Although our vision measures were all significantly correlated (ranging from Spearman correlation r_s = 0.26–0.91; P < 0.001), additional information may still be available from the use of two tests. The only combination that was better than plate 7 alone (κ_{0.5} = 0.93) was distance visual acuity and plate 7 (r_s = 0.76, κ_{0.5} = 0.94). Hence, there was no benefit in using combined tests over Arden plate 7 alone.

Detecting a particular disease

To allow comparison with some previous studies, we examined the ability of contrast sensitivity to discriminate those with particular ophthalmic disease diagnoses. Each of the three most prevalent principal diagnoses (Table 1) (age-related maculopathy, cataract, and glaucoma) was considered separately. Discrimination of those subjects with the ophthalmic disease diagnosis from all subjects with no diagnosed ophthalmic disease was evaluated. This is equivalent to the common study design of comparing selected subjects. Only the results for distance visual acuity and plate 7 are listed in Table 3, as other vision measures showed a relative discriminability similar to these results. There were no changes in tc, although κ_{0.5} varied for the different diseases with glaucoma being least well discriminated.

Because in a primary care practice we cannot expect to have a population composed only of subjects with and without a particular disease, we examined also the ability of contrast sensitivity to discriminate those with each of the three ophthalmic diseases from all other subjects. Again, only the results for distance visual acuity and plate 7 are listed (Table 4). All other vision measures showed similarly poor discrimination.

Discussion

This retrospective analysis of the study by Mitchell shows that contrast sensitivity, as measured using the Arden plate test, can better discriminate ophthalmic disease in an older population than Snellen visual acuity. This is contrary to the suggestions of some workers and to common clinical practice in which visual acuity is considered a more important measure of visual function. Plate 7 (6.4 cyc/deg) had the highest discriminability of the six test plates and was similar to the Modified Arden score. Although it is unlikely, the fixed order of presentation could have increased the discrimination with plate 7. Reliability and, hence, discriminability may improve with practice but may reduce with fatigue. Our expectation that the use of plate 7 after demonstration and practice with one of the earlier plates would show similar discriminability requires examination. No alternative method to Arden plate 7 alone—using a combination of visual acuity and contrast sensitivity, measuring contrast sensitivity only on those with better visual acuity, or assessing the interocular difference—was found to improve discriminability over that found using Arden plate 7 alone.

Kraemer and coworkers have described the advantages of the use of weighted kappa function, derived from the weighted kappa coefficient of association, κ_s, over traditional applications of signal detection approaches (κ_s accounts for misallocations and weights for relative costs of decisions). The outcomes from the weighted kappa function, QROC, and ROC analyses were not dissimilar as the prevalence was approximately 0.5 and we chose a weight-
often used inappropriate referent populations, failed to ap-
project the impact of many ophthalmic diseases can be re-
mitting, $r = 0.5$. Unfortunately, currently, there is little infor-
ning the discriminability of Bailey–Lovie visual acuity and letter
chart has been similarly criticized for poor reliability. 27  
contrast sensitivity at carefully selected letter sizes is war-
separate a recent article evaluating a range of vision tests and
manner, $\kappa_r$ also may be low when the prevalence of disease
leve l of agreement, and this occurs when the contingency table is highly skewed. 53 In a similar
weighted kappa function was easier to determine and to interpret.
acceptable weighting. The weighted kappa function
23. A revised psychometric method may improve reliability, 61 although a commercially available
An apparently paradoxical behavior of the kappa coefficient of agreement has been noted. 51–53 Under certain
This may be because of differences in the choice of decision criteria, disease prevalence (e.g., 0.15–0.64), and population characteristics (e.g., age range, method of sampling). Another difference from other studies using Arden plates 16,6 was the use by Mitchell18 of a novel psychometric method for Arden plate contrast sensitivity measurement.
The age-adjusted prevalence of diagnosed ophthalmic disease in the sample by Mitchell32 included all reported
diagnosis of an appropriate referent population is important.
Studies that have examined the ability of contrast sensitivity tests to screen for general ophthalmic disease within a
broad population have reported sensitivity and specificity lower than in this study. 3,6,7,9,11 This may be because of
discrimination between all other subjects. The optimal decision criterion ($t_c$) and the kappa coefficient ($\kappa_{0.5}$) at that
criterion are shown for Arden plate 7 contrast sensitivity and Snellen distance visual acuity.

### Table 4. Detecting a Particular Disease*

<table>
<thead>
<tr>
<th>Prevalence</th>
<th>$t_c$</th>
<th>$\kappa_{0.5}$</th>
<th>$t_c$</th>
<th>$\kappa_{0.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARM</td>
<td>0.15</td>
<td>-2</td>
<td>0.24</td>
<td>6/9</td>
</tr>
<tr>
<td>Cataract</td>
<td>0.13</td>
<td>-5</td>
<td>0.28</td>
<td>6/9</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>0.03</td>
<td>-2</td>
<td>0.05</td>
<td>6/60</td>
</tr>
</tbody>
</table>

ARM = age-related maculopathy.

* The vision measures could not discriminate well those with particular diagnosed ophthalmic disease in the right eye from all other subjects. The optimal decision criterion ($t_c$) and the kappa coefficient ($\kappa_{0.5}$) at that criterion are shown for Arden plate 7 contrast sensitivity and Snellen distance visual acuity.
when there is restricted access to or use of ophthalmic services. Whereas access restrictions occur frequently in developing countries, reductions in use can occur in the developed world. For example, the removal of the government provision of the “free sight test” in Britain caused a reduction in the number of persons seeking ophthalmic consultations. The natural long-term consequence will be more advanced ophthalmic disease at the time of detection. If those in need of ophthalmic care could be identified simply with a contrast sensitivity measure, for example in general medical practice or health clinics, there may be long-term savings.

In conclusion, contrast sensitivity has an important role in the primary care setting. Screening with contrast sensitivity was more efficient than with visual acuity. The Arden plate 7 had a positive predictive value of 0.96 and a negative predictive value of 0.97 for ophthalmic disease in the population studied. The merits of the weighted kappa function in the assessment of discriminability have been clearly demonstrated.

Appendix

The receiver-operating characteristics (ROC) function (Fig 1B) describes the performance of a test in terms of sensitivity and specificity over the range of test scores (decision criteria). The traditional measure used in ROC analysis of discriminability, *d*′, is an inappropriate measure of discriminability if the ROC function is not symmetric. A non-symmetric ROC function, which occurs when at least one of the population distributions is not normally distributed, is common in clinical settings. Highly skewed distributions often are found among groups of subjects with disease. A better measure of discriminability is the area under the ROC curve.

Although the optimal decision criterion (test score), *t*∗, in ROC analysis traditionally has been defined as the point on the ROC function closest to the point of perfect discrimination, *t*∗ should be corrected for prevalence. Thus, *t*∗ is the point on the ROC function intersected by the diagnosis line. Both these definitions of *t*∗ presume that the costs of decisions are equal. From a practical perspective, it often is not clear which is the optimal decision criterion. Finally, when the costs of decisions are unequal, *t*∗ can be defined as the point on the ROC function that has a slope *β* calculated as:

\[
\beta = \frac{1}{p} \cdot \frac{R_{fp} - R_{tn}}{R_{fp} - R_{tn}}
\]

where *p* is the prevalence of disease, *R* ∗, is the cost associated with a true-positive outcome, *R* ∗, is the cost associated with a false-negative outcome, etc. Ideally, the derivative of the ROC function should be determined. Because this is rarely possible, an alternative approach is to estimate the slope at each point by averaging the slope between the point of interest and points on either side. We approximated the slope at decision criterion *l*, *β* ∗, as:

\[
\beta_l = \frac{2}{\sqrt{p(1-p)}} \cdot \frac{y_{l+1} - y_l}{x_{l+1} - x_l} + \frac{y_l - y_{l-1}}{x_l - x_{l-1}}
\]

where *y* ∗ is the sensitivity and *x* ∗ is 1-specificity of criterion *l*.

As noted by Kraemer and Bloch a number of assumptions have been made in ROC analysis that, although appropriate in the electronic engineering context, are not appropriate in the clinical context. In particular, in the clinical setting there is the chance of misallocation of both presence of signal (disease) and of detector response (test result). The clinical experimenter can never be certain of the presence of the disease (signal), being reliant on some “gold standard.” An example is the imperfect agreement between practitioners on the assessment of the optic disc in glaucoma. In addition, a positive test outcome can occur because of events unrelated to the stimulus (disease). This is different from “noise.” Consequently, in clinical samples, sensitivity and specificity are not portable between clinical populations.

The weighted kappa coefficient of association, *κ* ∗, overcomes many of the problems associated with ROC analysis. *κ* ∗ is a generalised signal detection model of which ROC analysis is a subset (κ∗ is weighted according to the costs associated with the potential outcomes). Where *r* is the weighting factor and *r*′ = 1−*r*, *κ* ∗ can be defined as:

\[
κ_i = \frac{(n_{tp}n_{fn}) - (n_{fp}n_{tn})}{r(n_{tp} + n_{fn})(n_{fp} + n_{tn}) + r'(n_{fp} + n_{tn})(n_{tp} + n_{fn})}
\]

where *n* ∗ is the number with a true-positive outcome, *n* ∗ is the number with a false-negative outcome, etc. The value of *κ* ∗ ranges between 0 and 1 (like the more familiar correlation coefficient). The value of *r* also ranges between 0 and 1, being derived from the costs associated with each decision. Thus, there is a *κ* ∗ for each value of *r*, the appropriate *κ* ∗ depending on the relative costs associated with the decisions. When *r* = 1, maximal sensitivity will produce the optimal outcome, while if *r* = 0, maximal specificity will produce the optimal outcome. Unfortunately, in most contexts the relative costs are not known, in which case, *r* is usually fixed at 0.5 (κ∗,κ ∗ is known as the intraclass kappa coefficient). ROC analysis assumes *r* = 0.5.

Because *κ* ∗ and *κ* ∗ are rescaled versions of sensitivity and specificity, respectively, to describe discriminability, Kraemer recommended a new function that is known as the quality ROC (QROC). QROC reflects the quality of the decisions that can be made from the test outcome (Fig 1C). We found the weighted kappa functions, as recommended by Gilchrist, more useful in the evaluation of discriminability. Here, *κ* ∗ is plotted against each test decision criterion as shown in Figure 1D. The optimal criterion for discrimination, *t*∗, has the highest *κ* ∗ in the appropriate weighted kappa function. For example, in Figure 1D, where *r* = 0.5, the optimal decision criterion *t*∗ = 6/9, at which point *κ* ∗ = 0.61. This corresponds to a sensitivity of 0.74 and a specificity of 0.87 (Fig 1B, Table 2). The choice of the appropriate weighted kappa function depends on the costs
of incorrect outcomes. In the example given (Fig 1), the optimal decision criteria determined using ROC, QROC, and $\kappa_{0.5}$ were the same, but this is not always the case.

References


