RE: "META-ANALYSIS OF PAP TEST ACCURACY"

In their excellent meta-analysis, Fahey et al. (1) describe and attempt to explain variation in Papanicolaou (Pap) test screening accuracy for detection of cervical cancer and precancerous lesions. They have overlooked, though, important differences that arise when a screening test (such as the cervical smear) is compared with an imperfect reference test (one with sensitivity and/or specificity less than 100 percent). Fahey et al. focus on Pap test studies that used histologic diagnosis as the reference standard, but they admit that it is not perfect (1). Therefore, the effect of an imperfect reference test on the diagnostic accuracy of the Pap test is pertinent to their analysis, as I demonstrate below. In the remainder of this letter, I will refer to the test under evaluation, whether it is a screening or diagnostic test, as the diagnostic test, while the reference standard, also known as the gold standard or reference test, will be referred to as the reference test.

Several publications have demonstrated that, when a diagnostic test is evaluated for accuracy by comparison with an imperfect reference test, bias results in the estimation of diagnostic test sensitivity and specificity in a predictable direction (2-5). In general, when the reference test and diagnostic test errors are conditionally independent, the observed diagnostic test sensitivity and specificity underestimate the true values. As opposed to classic teaching, diagnostic test sensitivity and specificity vary with disease prevalence when the reference test is imperfect, such that observed sensitivity correlates positively while observed specificity correlates negatively with disease prevalence (2, 3). The observed diagnostic test sensitivity and specificity will not exceed the true values when conditional independence between diagnostic and reference test errors is present (5). My reanalysis of the data provided in the article by Fahey et al. indicates that a strong association is present between disease prevalence found in the studies included in their meta-analysis and Pap test sensitivity, specificity, and overall accuracy as reflected by the log odds ratio as defined by Fahey et al. (1). This analysis makes the reasonable assumption that the observed disease prevalence based on the imperfect reference test (histology) positively correlates with the true disease prevalence. Disease prevalence was calculated for each study considered in the meta-analysis as the proportion of subjects who tested positive by the reference test.

In table 1, as expected, when an imperfect reference test is used to evaluate a diagnostic test, the unweighted Pap test sensitivity increases somewhat with the higher disease prevalence category among the 59 studies included in the analysis, while the observed specificity decreases. The change in specificity by disease prevalence is of a larger magnitude compared with the change in sensitivity. The log odds ratio decreases with higher disease prevalence. A weighted linear regression analysis similar to that performed by Fahey et al. in their table 4 reveals a strong relation between diagnostic test accuracy and disease prevalence (table 2 below). For the regression models that considered either Pap test sensitivity or specificity as the dependent variable, the weight was defined as the inverse of the variance of the appropriate proportion (6). With the log odds ratio as the dependent variable, the weights used were those provided by Fahey et al. in their appendix 1. I could not examine the effects of other variables considered by Fahey et al. (independence of assessments, histologic threshold) in these regression models, because these values by study were not provided by their authors.

The weighted regression models also demonstrate the expected effect of an imperfect reference test on observed diagnostic test sensitivity and specificity. The former is positively correlated with disease prevalence, while the latter is negatively correlated. The log odds ratio is also negatively correlated with disease prevalence, which would be expected given the greater magnitude of the disease prevalence regression coefficient for specificity than sensitivity. These findings would be compatible with a reference standard that is more specific than sensitive, such that differences in true disease prevalence among studies lead to greater bias in subjects classified by the reference standard as disease free (in whom diagnostic test specificity is estimated) compared with diseased (in whom diagnostic test sensitivity is estimated). Disease prevalence showed the strongest statistical association with Pap test accuracy, as estimated by the log odds ratio. As opposed to the backward regression models performed by Fahey et al. that did not retain any statistically significant independent variables, the same technique when applied to the models in table 2 retained disease prevalence in the log odds ratio model and both disease prevalence and clinical use in the specificity model.

A solution to the problem of bias due to an imperfect reference standard is currently being sought, with a variety of approaches already having been taken to the problem (5, 7-9). Until this problem is overcome, it will be important for meta-analysts to consider the effects of reference test error on the variation in diagnostic test accuracy among studies. Not to do so may result in missing an important source of such variation, as demonstrated in this example.

REFERENCES


TABLE 1. Reanalysis of data from Fahey et al.* Indicating a strong association between disease prevalence in their studies and Papanicolaou test sensitivity, specificity, and overall accuracy

<table>
<thead>
<tr>
<th>Disease prevalence category</th>
<th>No. of studies</th>
<th>Mean sensitivity</th>
<th>Mean specificity</th>
<th>Mean log odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01-0.33</td>
<td>14</td>
<td>0.51</td>
<td>0.81</td>
<td>1.94</td>
</tr>
<tr>
<td>0.34-0.66</td>
<td>23</td>
<td>0.64</td>
<td>0.67</td>
<td>1.51</td>
</tr>
<tr>
<td>0.67-0.99</td>
<td>22</td>
<td>0.64</td>
<td>0.59</td>
<td>1.29</td>
</tr>
</tbody>
</table>

TABLE 1. Variation in Papanicolaou test accuracy (log odds ratio) due to study characteristics

<table>
<thead>
<tr>
<th>Predictors (R² = 0.22)</th>
<th>Beta*</th>
<th>95% CI†</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease prevalence</td>
<td>-1.14</td>
<td>-2.02 to -0.26</td>
<td>0.01</td>
</tr>
<tr>
<td>Clinical use</td>
<td>0.71</td>
<td>0.00 to 1.42</td>
<td>0.05</td>
</tr>
<tr>
<td>Histologic threshold</td>
<td>-0.09</td>
<td>-0.49 to 0.31</td>
<td>0.64</td>
</tr>
<tr>
<td>Independence of assessments</td>
<td>0.57</td>
<td>-0.15 to 1.29</td>
<td>0.12</td>
</tr>
<tr>
<td>Year of publication</td>
<td>-0.04</td>
<td>-0.10 to 0.02</td>
<td>0.23</td>
</tr>
</tbody>
</table>

* Beta denotes the partial regression coefficient for the full model. Coding clinical use, screening = 0, follow-up = 1, histologic threshold, cervical intraepithelial neoplasia grade 1 = 0, cervical intraepithelial neoplasia grade 2 = 1, independence of assessments, unblinded = 0, blinded = 1.
† CI, confidence interval.

REFERENCES

Michael T. Fahey
Epidemiology Unit
Liverpool Hospital
Liverpool, NSW 2170
Australia

Les Irwig
Department of Public Health
University of Sydney
Sydney, NSW 2006
Australia

Petra Macaskill
Epidemiology and Surveillance Unit
Worksafe Australia
GPO Box 58
Sydney, NSW 2001
Australia

Am J Epidemiol Vol. 143, No. 4, 1996