Journal of Biometrics & Biostatistics

Cotton et al., J Biom Biostat 2015, 6:1
DOI: 10.4172/2155-6180.1000217

Research Article Open Access

Using Available Information in the Assessment of Diagnostic Protocols

Cecilia A Cotton*, Oana Danila, Stefan H Steiner, Daniel Severn and R Jock MacKay

Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave West, Waterloo, ON, Canada N2L 3G1

Abstract

A new binary screening or diagnostic test may be combined sequentially with an existing test using either a believe the positive or believe the negative protocol. Interest then lies in estimating the properties of the new combined protocol and in comparing the new protocol with the existing test via sensitivity, specificity, or likelihood ratios that capture the trade-off between sensitivity and specificity. We consider a paired assessment study with complete verification via a gold standard. Our goal is to quantify the gain in precision for the estimators of the sensitivity, specificity and the ratio of likelihood ratios in protocols when baseline information on the performance of the existing test is available. We find maximum likelihood estimators of the quantities of interest and derive their asymptotic standard deviations. The methods are illustrated using previously published mammography and ultrasound test results from a cohort of symptomatic women. We find that incorporating baseline information has a large impact on the precision of the estimator for the specificity of the believe the positive protocol and of the sensitivity of the believe the negative protocol. Including available baseline information can improve the precision of estimators of the sensitivity, specificity, and the ratio of likelihood ratios and/or reduce the number of subjects needed in an assessment study to evaluate the protocol.

Keywords: Binary diagnostic test; Sensitivity; Specificity; Likelihood ratios; Combined tests; Baseline data

Introduction

In screening and diagnostic testing, it is common for researchers to consider how a new diagnostic test might be combined sequentially with an existing test to either improve the overall diagnostic performance or to reduce the overall cost while maintaining performance. If we assume that the tests produce binary results there are two protocol choices. A *believe the positive* protocol produces a positive result for all combinations of the existing and new test results except when both tests are negative while a *believe the negative* protocol produces a positive result only when both tests are positive. In practice the tests may be applied in sequence so we refer to an *add-on* plan as one in which the existing test is used first with the new test applied to only a subset of subjects depending on the protocol in use. In a *triage* plan, the new test is used on all subjects and the results determine who receives the existing test. Therefore, with two binary tests, a total of four sequential diagnostic protocols are possible.

Sequential protocols are used in a wide variety of settings including computed tomography as an add-on test to pelvic ultra sonography in a believe the positive protocol for the diagnosis of suspected appendicitis in children in Peña *et al.* [1], and high-risk human papillama virus testing as a triage test to cytological screening in a believe the negative protocol for precancerous cervical lesions in Kotaniemi-Talonen *et al.* [2]. Although we refer to existing and new tests either could itself be a protocol combining multiple tests as long as the final result is dichotomous. For example, Gyselaers *et al.* [3] designed a screening protocol for trisomy 21 detection in the first trimester using a combination of age and serum test results as a triage to identify intermediate risk cases for advanced ultrasound scanning.

Our purpose is to demonstrate the value, expressed in terms of improved precision of important estimators, of using available information on the performance of the existing test, in combination with data from an assessment study of the existing and new tests to assess the properties of a combined protocol and compare the protocol to the existing test.

The available information or baseline data takes the form of a previous case-control study used to estimate the sensitivity and specificity of the existing test [4,5]. Initially, we confine ourselves to the setting where the assessment study is also a case-control study using a paired design (i.e. both tests are applied to all subjects). We assume that verification via a gold standard is available for both the assessment study and the baseline data. Later we discuss the use of cohort studies, where the numbers of diseased and non-diseased subjects are not fixed, and non-paired sequential assessment studies, where the new test is given to only a subset of subjects.

We make a key assumption that the diagnostic properties (sensitivity and specificity) of the existing test are the same in the baseline and assessment studies. Practically, this means selecting a source of baseline data with a population and test implementation as similar as possible to the assessment study. Depending on the disease under study, important population factors might include age- and sex-distributions and disease strain or severity in the cases. We derive a formal test of this hypothesis. If we reject the hypothesis then it is likely not appropriate to incorporate the baseline data into the analysis and either a new source of baseline data should be found or a standard analysis of the assessment study should be conducted. However, if we fail to reject the hypothesis, we will proceed with incorporating the available baseline data into the analysis of the combined protocol.

To assess the properties of the new protocol, we use maximum likelihood estimation for the sensitivity and specificity and asymptotic properties of the likelihood to find approximate standard errors. The comparison of the new protocol versus the existing test is made via likelihood ratios. These capture the trade-off between sensitivity and specificity of the new protocol versus the existing test [6-8]. Informally,

***Corresponding author:** Cecilia A Cotton, Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave West, Waterloo, ON, Canada N2L 3G1, Tel:+1 519-888-4567; E-mail: ccotton@uwaterloo.ca

Received December 03, 2014; **Accepted** February 02, 2015; **Published** February 20, 2015

Citation: Cotton CA, Danila O, Steiner SH, Severn D, MacKay RJ (2015) Using Available Information in the Assessment of Diagnostic Protocols. J Biom Biostat 6: 217. doi:10.4172/2155-6180.1000217

the positive (negative) likelihood ratio represents how many times more often a subject with the disease is expected to have a positive (negative) test result than one without the disease. The comparison between likelihood ratios is equivalent to comparing the predictive values of the two tests, when applied to the same population of subjects.

Likelihood ratios have previously been used by Nofuentes and del Castillo [9] to compare and test for differences between two binary tests and by Macaskill *et al.* [7] to compare a combined test with one of its components. Muwonge *et al.* [10] considered the sensitivities, specificities, and likelihood ratios for combinations of screening tests for cervical cancer. Bossuyt *et al.* [11] discussed the various ways of combining a new with an existing diagnostic test and Hayen *et al.* [8] discussed estimation of the likelihood ratios for a full paired design. Ahmed *et al.* [12] and Ahmed *et al.* [13] considered three logic rules for combining two tests based on continuous markers and compared them in terms of the false positive rate, the maximum receiver operating characteristic curve, and overall cost. While sequential tests are relatively common in medicine, we are not aware of previous work on incorporating baseline data. The use of baseline data is more common in industrial applications [14,15].

The remainder of the article is organized as follows. In the Methods section we define notation and formulate the sensitivity and specificity of the combined protocols and the likelihood ratios used to compare the believe the positive or believe the negative protocol versus the existing test. We also derive the relevant asymptotic variances and discuss estimation. In the Application section we apply the methods to an example of the use of mammography and sonography/ultrasound for breast cancer detection in symptomatic women. Next we conduct a factorial experiment to study the gains in precision across a wide range of sensitivities, specificities and study sizes. Finally, we conclude with a discussion.

Methods

The results of a paired case-control assessment study of two binary diagnostic tests with verification via a gold standard can be summarized in two 2×2 tables, one for diseased subjects and one for non-diseased subjects. Let *E* and *N* be binary random variables representing the result (1 for a positive result, 0 for a negative result) of the existing and new tests, respectively. Let *D* be a binary random variable representing the disease status (1 for the diseased, 0 for the non-diseased). We denote the cell frequencies for the assessment study by $_{en}$ with *e,n,d*=0,1. Let n_1 be the total number of diseased subjects in the assessment study $(n_1 = n^1_{00} + n^1_{01} + n^1_{10} + n^1_{11})$, n_0 be the total non-diseased subjects, and $n = n_1 + n_0$ be the total sample size.

We define $\pi^d_{en} = \Pr[E = e, N = n \mid D = d]$ to be the joint probability of the results of the existing and new tests conditional on disease status. Using the convention of a + indicating marginalization or summation over that subscript, the sensitivity and specificity of the existing test from the paired assessment study are given by:

$$sens_E = Pr(E = 1 \mid D = 1) = \pi^1_{1+} = 1 - \pi^1_{0+} \tag{1}$$

$$spec_E = Pr(E = 0 \mid D = 0) = 1 - \pi^0_{1+} = \pi^0_{0+} \tag{2}$$

Additionally, we assume we have available baseline information about the performance of the existing test. We denote these data by m^d_e, *e*, *d*=0,1. The total number of diseased subjects in the baseline is denoted by m_1, the number of non-diseased subjects by m_0, and the total sample size for the baseline study by $m=m_1 + m_0$. An example of data of this type is given in Table 1 which will be discussed in detail in the Application section. Before we incorporate the available information in the estimation of the properties of a believe the positive or believe the negative protocol we recommend testing the equivalence of the sensitivity and specificity estimates of the existing test in the baseline and assessment studies. A likelihood ratio test for this purpose is given in Appendix A of the Supplementary Materials. Assuming we do not reject the equivalence hypothesis, we now develop the methodology necessary to incorporate the available information into the estimation of the properties of a believe the positive or believe the negative protocol.

Paired assessment study data								Baseline data		
Breast cancer cases				Non breast cancer controls						
	US+	US-	Total		US+	US-	Total		Cases	Non-Cases
Mx+	148	34	182	Mx+	8	21	29	Mx+	98	1062
Mx-	48	10	58	Mx-	20	184	204	Mx-	40	9208
Total	196	44	240		28	205	233		138	10270

Table 1: Results of mammogram (Mx) and ultrasound (US) screening from the parallel assessment study of Houssami et al. [16] and mammogram baseline data from Kavanagh et al. [17].

Believe the positive protocol

When two tests are combined there is always a trade-off between the sensitivity and specificity of the new protocol and the existing test. Believe the positive protocols are used to reduce the number of false negatives and therefore increase the sensitivity over the existing test. Since a positive result on either the existing or new test yields a positive protocol result we will denote the protocol as $E \cup N$. The conditional probabilities of the existing test and the believe the positive protocol results, given the true disease status, can be expressed in terms of the conditional probabilities corresponding to the cell frequencies in a paired assessment study. For *D*=1, these conditional probabilities are:

$$\begin{aligned} \Pr\left(E = 1, E \cup N = 1 \mid D = 1\right) &= \Pr\left(E = 1 \mid D = 1\right) = \pi^1_{11} + \pi^1_{10} = \pi^1_{1+} \\ \Pr\left(E = 0, E \cup N = 1 \mid D = 1\right) &= \Pr\left(E = 0, N = 1 \mid \mathrm{D} = 1\right) = \pi^1_{01} \\ \Pr\left(E = 0, E \cup N = 0 \mid D = 1\right) &= \Pr\left(E = 0, N = 0 \mid \mathrm{D} = 1\right) = \pi^1_{00} \end{aligned} \tag{3}$$

and similarly for *D*=0. Note that we can write $\pi^1_{00} = 1 - \pi^1_{1+} - \pi^1_{01}$ and $\pi^0_{00} = 1 - \pi^0_{1+} - \pi^0_{01}$.

The sensitivity and specicity of the believe the positive protocol can be expressed as:

$$sens_{E \cup N} = Pr(E \cup N = 1 \mid D = 1) = \pi^1_{1+} + \pi^1_{01} \tag{4}$$

$$spec_{E \cup N} = Pr(E \cup N = 0 \mid D = 0) = 1 - \pi^0_{1+} - \pi^0_{01} \tag{5}$$

Through comparisons to Equations (1) and (2), note that $sens_{E \cup N} \geq sens_E$ and $spec_{E \cup N} \leq spec_E$. Therefore, since higher sensitivity and specificity are desirable, there is a trade-off as the believe the positive protocol will always have a sensitivity at least as large as the existing test, and a specificity no greater than the existing test.

We can also compare the existing test and the new protocol in terms of their positive predictive values (PV^+) and negative predictive values (PV^-)

$$PV^+_E = Pr\left(D = 1 \mid E = 1\right); \qquad PV^-_E = Pr\left(D = 0 \mid E = 0\right)$$
$$PV^+_{E \cup N} = Pr(D = 1 \mid E \cup N = 1); \quad PV^-_{E \cup N} = Pr(D = 0 \mid E \cup N = 0)$$

As the predictive values are post-test probabilities of disease (non-disease) among the population of subjects who tested positive (negative), they are important from a clinical point of view. The trade-off between specificity and sensitivity does not necessarily translate into a trade-off between the two predictive values. It is possible for the believe the positive protocol to have higher positive and negative predictive values than the existing test. However, it is well known that the predictive values of a test depend on the accuracy of the test and the disease prevalence [7]. Therefore, a direct comparison between the predictive values of two tests is not recommended, as the results of the comparison cannot be extended beyond the population on which the assessment study was conducted.

Macaskill *et al.* [7] and Hayen *et al.* [8] propose comparing the new protocol with the existing test by looking at the logarithm of the ratio of their corresponding positive and negative likelihood ratios. The likelihood ratios of the two tests can be expressed in terms of the conditional probabilities defined in Equation (3) as follows

$$LR_E^+ = \frac{sens_E}{1-spec_E} = \frac{\pi_{1+}^1}{\pi_{1+}^0}; \quad \mathrm{LR}_E^- = \frac{1-sens_E}{spec_E} = \frac{1-\pi_{1+}^1}{1-\pi_{1+}^0}$$

$$LR_{E\cup N}^+ = \frac{\pi_{1+}^1+\pi_{01}^1}{\pi_{1+}^0+\pi_{01}^0}; \qquad \mathrm{LR}_{E\cup N}^- = \frac{1-\pi_{1+}^1-\pi_{01}^1}{1-\pi_{1+}^0-\pi_{01}^0}$$

Marshall [6] shows that comparing the predictive values of two tests is equivalent to comparing the likelihood ratios of these tests, for fixed prevalence. That is, for fixed prevalence,

$$PV^+{}_{E\cup N} \geq PV_E^+ \text{ if and only if } LR^+{}_{E\cup N} \geq LR_E^+$$

and

$$PV^-{}_{E\cup N} \geq PV_E^- \text{ if and only if } LR^-{}_{E\cup N} \leq LR_E^-$$

However, since the likelihood ratios do not depend on the disease prevalence we can compare the estimates of the likelihood ratios from the assessment study directly. This comparison translates into a comparison between the predictive values of the tests when applied to the same population.

The expressions for the logarithm of the ratios of the positive and negative likelihood ratios for the believe the positive protocol, denoted here $\phi_{E\cup N}^+$ and $\phi_{E\cup N}^-$, respectively, in terms of the conditional probabilities, are as follows:

$$\phi_{E\cup N}^+ = \ln\left(\frac{LR^+_{E\cup N}}{LR_E^+}\right) = [\ln(\pi_{1+}^1+\pi_{01}^1) - \ln(\pi_{1+}^1)] - [\ln(\pi_{1+}^0+\pi_{01}^0) - \ln(\pi_{1+}^0)] \qquad (6)$$

$$\phi_{E\cup N}^- = \ln\left(\frac{LR^-_{E\cup N}}{LR_E^-}\right) = [\ln(1\text{-}\pi_{1+}^1+\pi_{01}^1) - \ln(1\text{-}\pi_{1+}^1)] - [\ln(1\text{-}\pi_{1+}^0+\pi_{01}^0) - \ln(1\text{-}\pi_{1+}^0)] \qquad (7)$$

If the believe the positive protocol is better than the existing test in terms of both the positive and negative predictive values then $\phi_{E\cup N}^+ > 0$ and $\phi_{E\cup N}^- < 0$. However, it might be that the new protocol is better than the existing one in terms of the positive predictive value, but worse in terms of the negative predictive value, or the other way around.

Analysis for the believe the positive protocol

In this section, our goal is to quantify the gain in precision for the estimators of sensitivity and specificity of the believe the positive protocol, and of the log-ratios for both likelihood ratios, i.e., $\phi_{E\cup N}^+$ and $\phi_{E\cup N}^-$ that comes from incorporating the available baseline information into the estimation procedure.

Assuming independence between the baseline and assessment data, we can write the likelihood function as $L \propto L_b \times L_s$ where L_b is the baseline likelihood

$$L_b \propto (\pi_{1+}^1)^{m_1^1}(1-\pi_{1+}^1)^{m_1-m_1^1} \times (\pi_{1+}^0)^{m_1^0}(1-\pi_{1+}^0)^{m_0-m_1^0} \qquad (8)$$

and L_s is the assessment study likelihood

$$L_s \propto (\pi_{1+}^1)^{n_{1+}^1}(\pi_{01}^1)^{n_{01}^1}(1-\pi_{1+}^1-\pi_{01}^1)^{\mathrm{n}_{00}^1} \times (\pi_{1+}^0)^{\mathrm{n}_{1+}^0}(\pi_{01}^0)^{\mathrm{n}_{01}^0}(1-\pi_{1+}^0-\pi_{01}^0)^{\mathrm{n}_{00}^0} \qquad (9)$$

Note that here we assume that the sensitivity (π_{1+}^1) and specificity ($1-\pi_{1+}^0$) of the existing test are the same in the baseline and assessment studies. This assumption should be tested as outlined in Appendix A of the Supplementary Materials. Using the combined likelihood function, the maximum likelihood estimates (MLEs) of the four conditional probabilities are

$$\hat{\pi}_{1+}^D = \frac{m_1^D + n_{1+}^D}{m_D + n_D}; \hat{\pi}_{01}^D = \frac{(m_D + n_D - m_1^D - n_{1+}^D)n_{01}^D}{(m_D+n_D)(n_D - n_{1+}^D)}, \text{ for } D = 0,1. \qquad (10)$$

The MLEs of the sensitivity and specificity of the believe the positive protocol and of the log-ratios of the likelihood ratios are obtained by substituting the MLEs given in Equation (10) into Equations (4-7).

The asymptotic variances of the estimators of $sens_{E\cup N}$, $spec_{E\cup N}$, $\phi_{E\cup N}^+$ and $\phi_{E\cup N}^-$ are found using the delta method (see the Appendix B of the Supplementary Materials for complete derivations). Given the numbers of diseased and non-diseased subjects in the assessment study and the baseline data, the asymptotic variances for the estimators of $sens_{E\cup N}$ and $spec_{E\cup N}$ are

$$\mathrm{var}[\widehat{sens}_{E\cup N}] = \frac{\pi_{1+}^1(1-\pi_{1+}^1) - 2\pi_{1+}^1\pi_{01}^1}{n_1+m_1} + \frac{\pi_{01}^1[n_1\pi_{1+}^1\pi_{01}^1 + (n_1+m_1)(1-\pi_{1+}^1-\pi_{01}^1)]}{n_1(n_1+m_1)(1-\pi_{1+}^1)} \qquad (11)$$

$$\mathrm{var}[\widehat{spec}_{E\cup N}] = \frac{\pi_{1+}^0(1-\pi_{1+}^0) - 2\pi_{1+}^0\pi_{01}^0}{n_0+m_0} + \frac{\pi_{01}^0[n_0\pi_{1+}^0\pi_{01}^0 + (n_0+m_0)(1-\pi_{1+}^0-\pi_{01}^0)]}{n_0(n_0+m_0)(1-\pi_{1+}^0)} \qquad (12)$$

For the estimators of $\phi_{E\cup N}^+$, and $\phi_{E\cup N}^-$

$$\mathrm{var}[\hat{\phi}_{E\cup N}^+] = \frac{\pi_{01}^1[m_1\pi_{1+}^1 + (\pi_{1+}^1+\pi_{01}^1)(n_1 - m_1\pi_{1+}^1 - n_1\pi_{1+}^1)]}{n_1(n_1+m_1)\pi_{1+}^1(\pi_{1+}^1+\pi_{01}^1)^2(1-\pi_{1+}^1)} + \frac{\pi_{01}^0[m_0\pi_{1+}^0 + (\pi_{1+}^0+\pi_{01}^0)(n_0 - m_0\pi_{1+}^0 - n_0\pi_{1+}^0)]}{n_0(n_0+m_0)\pi_{1+}^0(\pi_{1+}^0+\pi_{01}^0)^2(1-\pi_{1+}^0)} \qquad (13)$$

$$\mathrm{var}[\hat{\phi}_{E\cup N}^-] = \frac{\pi_{01}^1}{n_1(1-\pi_{1+}^1-\pi_{01}^1)(1-\pi_{1+}^1)} + \frac{\pi_{01}^0}{n_0(1-\pi_{1+}^0-\pi_{01}^0)(1-\pi_{1+}^0)} \qquad (14)$$

If no baseline data are available, the MLEs and expressions for the asymptotic variances can be obtained by substituting $m_1 = m_0 = 0$ in Equation (10) and Equations (11-14), respectively. We note here that even when it is available, the MLE for $\phi_{E\cup N}^-$ and the expression for its asymptotic variance given by Equation (14) do not depend on the baseline data. However, the asymptotic variances of, $\widehat{sens}_{E\cup N}$ $\widehat{spec}_{E\cup N}$ and $\hat{\phi}_{E\cup N}^+$ are always reduced by including the baseline data.

Equations (11-14) are used to derive the standard errors for the estimates of sensitivity and specificity of the believe the positive protocol, and for the log-ratios of the positive and negative likelihood ratios of the two tests, by replacing the values of π_{1+}^D, and π_{01}^D, D=0,1, with their corresponding MLEs, and then taking the square root. In cases where the sample size for the assessment study is small or the counts for some n_{en}^d are very small, other methods such as bootstrapping are recomended for obtaining the standard errors [7].

Believe the negative protocol

Believe the negative protocols are used to reduce the number of false positives and therefore increases the specificity over the existing test. We will denote the protocol as $E \cap N$. The conditional probabilities, given the true disease status, of the existing test and the believe the negative protocol results are:

$$\begin{aligned} &Pr(E=0, E \bigcap N=0 \mid D=1) = Pr(\mathrm{E}=0 \mid \mathrm{D}=1) \quad = \pi_{00}^1 + \pi_{01}^1 = \pi_{0+}^1 \\ &Pr(E=1, E \bigcap N=1 \mid D=1) = Pr(\mathrm{E}=1, N=1 \mid \mathrm{D}=1) = \pi_{11}^1 \\ &Pr(E=1, E \bigcap N=0 \mid D=1) = Pr(\mathrm{E}=1, N=0 \mid \mathrm{D}=1) = \pi_{10}^1 \end{aligned} \tag{15}$$

and similarly for D=0. Note that we can write $\pi_{11}^1 = 1 - \pi_{1+}^1 - \pi_{10}^1$ and $\pi_{11}^0 = 1 - \pi_{1+}^0 - \pi_{10}^0$.

The sensitivity and specificity of the believe the negative protocol can be expressed as:

$$sens_{E \cap N} = Pr(E \bigcap N = 1 \mid D=1) = \pi_{11}^1 = 1 - \pi_{0+}^1 - \pi_{10}^1 \tag{16}$$

$$spec_{E \cap N} = Pr(E \bigcap N = 0 \mid D=0) = \pi_{0+}^0 + \pi_{10}^0 \tag{17}$$

The positive and negative likelihood ratios for the two tests are given by

$$LR_E^+ = \frac{sens_E}{1 - spec_E} = \frac{1-\pi_{0+}^1}{1-\pi_{0+}^0}; LR_E^- = \frac{1 - sens_E}{spec_E} = \frac{\pi_{0+}^1}{\pi_{0+}^0}$$

$$LR_{E \cap N}^+ = \frac{1 - \pi_{0+}^1 - \pi_{10}^1}{\pi_{0+}^0 + \pi_{10}^0}; LR_{E \cap N}^- = \frac{\pi_{0+}^1 + \pi_{10}^1}{1 - \pi_{0+}^0 + \pi_{10}^0}$$

Note that $sens_{E \cap N} \leq sens_E$ and $spec_{E \cap N} \geq spec_E$. Therefore, as in the case of a believe the positive protocol, there is a trade-off between sensitivity and specificity, but this time the new protocol improves the specificity, but decreases the sensitivity. We can compare the performance of the existing test to the new protocol in terms of the logarithm of the ratio of their positive and negative likelihood ratios

$$\phi_{E \cap N}^+ = \ln\left(\frac{LR_{E \cap N}^+}{LR_E^+}\right) = [\ln(1 - \pi_{0+}^1 - \pi_{10}^1) - \ln(1 - \pi_{0+}^1)] - [\ln(\pi_{0+}^0 + \pi_{10}^0) - \ln(1 - \pi_{0+}^0)] \tag{18}$$

$$\phi_{E \cap N}^- = \ln\left(\frac{LR_{E \cap N}^-}{LR_E^-}\right) = [\ln(\pi_{0+}^1 + \pi_{10}^1) - \ln(\pi_{0+}^1)] - [\ln(1 - \pi_{0+}^0 - \pi_{10}^0) - \ln(\pi_{0+}^0)] \tag{19}$$

For fixed prevalence, the comparison between the likelihood ratios translates into a comparison between predictive values, as explained for the believe the positive protocol. In practice, there may be situations where the believe the negative protocol is better than the existing test in terms of the positive predictive values and worse in terms of the negative predictive value, or vice versa.

Analysis for the believe the negative protocol

The analysis of the believe the negative protocol mirrors the analysis of the believe the positive protocol. We reparameterize the likelihoods for the baseline and assessment study in terms of the four parameters π_{0+}^D, π_{10}^D, D=0,1 to obtain

$$L_b \propto (1 - \pi_{0+}^1)^{m_1^1} (\pi_{0+}^1)^{m_1 - m_1^1} \times (1 - \pi_{0+}^0)^{m_1^0} (\pi_{0+}^0)^{m_0 - m_1^0}$$

$$L_s \propto (\pi_{0+}^1)^{n_{0+}^1} (\pi_{10}^1)^{n_{10}^1} (1 - \pi_{0+}^1 - \pi_{10}^1)^{n_{11}^1} \times (\pi_{0+}^0)^{n_{0+}^0} (\pi_{10}^0)^{n_{10}^0} (1 - \pi_{0+}^0 - \pi_{10}^0)^{n_{11}^0}$$

The MLEs of the four conditional probabilities are

$$\hat{\pi}_{0+}^D = \frac{m_0^D + n_{0+}^D}{m_D + n_D}; \hat{\pi}_{10}^D = \frac{(m_D + n_D - m_0^D - n_{0+}^D)\mathrm{n}_{10}^D}{(m_D + n_D)(n_D - n_{0+}^D)} \text{ for } D = 0,1. \tag{20}$$

The MLEs of the sensitivity and specificity of the believe the negative protocol and of the log-ratios of the likelihood ratios are obtained by substituting the MLEs given in Equation (20) into Equations (16-19).

Conditioning on the numbers of diseased and non diseased subjects in the assessment study and the baseline data, the expressions for the asymptotic variance for the estimations of $sens_{E \cap N}$, $spec_{E \cap N}$, $\phi_{E \cap N}^+$, and $\phi_{E \cap N}^-$ are

$$\mathrm{var}[\widehat{sens}_{E \cap N}] = \frac{\pi_{0+}^1(1 - \pi_{0+}^1) - 2\pi_{0+}^1\pi_{10}^1}{n_1 + m_1} + \frac{\pi_{10}^1[n_1\pi_{0+}^1\pi_{10}^1 + (n_1 + m_1)(1 - \pi_{0+}^1 - \pi_{10}^1)]}{n_1(n_1 + m_1)(1 - \pi_{0+}^1)} \tag{21}$$

$$\mathrm{var}[\widehat{spec}_{E \cap N}] = \frac{\pi_{0+}^0(1 - \pi_{0+}^0) - 2\pi_{0+}^0\pi_{10}^0}{n_0 + m_0} + \frac{\pi_{10}^0[n_0\pi_{0+}^0\pi_{10}^0 + (n_0 + m_0)(1 - \pi_{0+}^0 - \pi_{10}^0)]}{n_0(n_0 + m_0)(1 - \pi_{0+}^0)} \tag{22}$$

$$\mathrm{var}[\hat{\phi}_{E \cap N}^+] = \frac{\pi_{10}^1}{n_1(1 - \pi_{0+}^1 - \pi_{10}^1)(1 - \pi_{0+}^1)} + \frac{\pi_{10}^0}{n_0(1 - \pi_{0+}^0 - \pi_{10}^0)(1 - \pi_{0+}^0)} \tag{23}$$

$$\mathrm{var}[\hat{\phi}_{E \cap N}^-] = \frac{\pi_{10}^1[m_1\pi_{0+}^1 + (\pi_{0+}^1 + \pi_{10}^1)(n_1 - m_1\pi_{0+}^1 - n_1\pi_{0+}^1)]}{n_1(n_1 + m_1)\pi_{0+}^1(\pi_{0+}^1 + \pi_{10}^1)^2(1 - \pi_{0+}^1)} + \frac{\pi_{10}^0[m_0\pi_{0+}^0 + (\pi_{0+}^0 + \pi_{10}^0)(n_0 - m_0\pi_{0+}^0 - n_0\pi_{0+}^0)]}{n_0(n_0 + m_0)\pi_{0+}^0(\pi_{0+}^0 + \pi_{10}^0)^2(1 - \pi_{0+}^0)} \tag{24}$$

There is a clear symmetry between these results and those for the believe the positive protocol in Equations (11-14). If baseline data are not available, the MLEs can be obtained by substituting $m_1 = m_0 = 0$ into the relevant equations. For the believe the negative protocol the MLE for $\phi_{E \cap N}^+$ and the expression for its asymptotic variance given by Equation (23) do not depend on the baseline data.

Additional Considerations

Add-on and triage plans: The believe the positive and believe the negative protocols determine the rule for combining the results of the two tests. In practice the protocols can be implemented using either an add-on or triage plan. In an add-on plan the existing test is first given to all subjects. Follow-up testing with the new test is conducted only on those with an initial negative results for believe the positive or initial positive result for believe the negative. For a triage plan, the new test is used first and the results determine who will receive the existing test. The choice of plan and protocol comes down to scientific and practical considerations. For instance if the new test is invasive or expensive it may best be used as an add-on in a believe the negative protocol to improve sensitivity over the use of the existing test alone. This combination means that only subjects likely to be truly diseased (on the basis of a positive result from the existing test) are given the new test. The diagnostic properties of the protocols do not depend on whether they will be used with an add-on or triage plan.

Case-control versus cohort studies: The expressions for the MLEs in Equations (10) and (20) and the asymptotic variances in Equations (11-14) and Equations (21-24) apply to settings where the numbers of diseased, n_1 and m_1 and non-diseased, n_0 and m_0, in the assessment and baseline studies are fixed, as in a case-control design. If the assessment study is designed as a cohort study, these expressions still apply, with n_1 replaced by $n\theta$ and n_0 by $n(1-\theta)$, where θ is the prevalence of disease and n is the total sample size. Similar adjustments can be made to m_1 and m_0 if the baseline data come from a cohort study.

Paired versus sequential assessment studies: The analysis methodology given above assumes that a paired assessment study is used. If it is known a priori whether the new and existing tests will be combined using a believe the positive or believe the negative protocol a sequential assessment study can be used instead. We still assume full verification via a gold standard. First, the existing test should be applied to all subjects. This is necessary so that we can test for equivalence of the diagnostic properties of the existing test across the baseline data and assessment study. Following that, the new test should be applied to only those subjects with an initial negative result for a believe the positive protocol or initial positive result for a believe the negative protocol. This means we do not observed the full 2×2 data tables. However, the MLEs and asymptotic variance expressions under a believe the positive protocol in Equations (10-14) only depend on n^D_{1+}, n^D_{01} and n_D so it is not necessary to separately observe (n^D_{11}, n^D_{10}), D=1,0. The same holds for a believe the negative protocol where it is not necessary to separately observe (n^D_{01}, n^D_{00}), D=1,0.

Application

As an example, we consider breast cancer detection in symptomatic women. In a case-control study, Houssami *et al.* [16] examined the mammography (Mx) and sonography/ultrasound (US) test results of 240 women shown to have breast cancer and 233 age-matched controls. Both cases and controls were women aged 25 to 55 years who had been referred to testing due to the presence of symptoms (a palpable lump, pain or localized discomfort, etc.). The controls were selected among those who were not diagnosed with breast cancer in the two years following their assessment. The images were interpreted by two radiologists blinded to the subject's cancer status. Since both the cases and controls were tested using both modalities, the paired data are available in Table 1. A complete discussion of the study design and inclusion and exclusion criteria is available in Houssami *et al.* [16]. While the original purpose of the study was to investigate the age-specific sensitivity and specificity of the tests, we will use the data as an illustration to evaluate whether ultrasound is useful when combined with mammography in a believe the positive or believe the negative protocol. Based on the paired data in Table 1, $\widehat{sens}_{\text{Mx}}$=75.8%, $\widehat{spec}_{\text{Mx}}$ =87.6%, $\widehat{sens}_{\text{US}}$ =81.6%, and $\widehat{spec}_{\text{US}}$ =88.0%. To illustrate, we treat mammography as the existing test and ultrasound as a new test. Among the cases and controls, 34.1% and 17.6%, respectively, of the women had discrepant mammography and ultrasound test results.

In order to proceed with the methods described in the previous section, we need baseline data on the performance of mammography in the given population. We use the data provided in Kavanagh *et al.* [17] for this purpose (see Table 1-Baseline data) which provide estimates of the sensitivity and specificity of mammography in a similar symptomatic population. We consider the subpopulation of women with any breast symptoms and estimate the sensitivity and specificity of mammography to be 71.0% and 89.7%, respectively. Both studies were conducted among Australian women in the mid 1990's. However, the population in the baseline study was slightly older which may cause an inflation of the sensitivity estimate. However, the likelihood ratio test derived in Appendix A of the Supplementary Materials gave no evidence (p-value=0.35) to reject the hypothesis that the two accuracy measures are common for the assessment study and baseline population We consider two diagnostic protocols:

1. Mx ∪ US, a believe the positive protocol with US as an add-on to Mx. This protocol will have better sensitivity than Mx alone and uses US to catch potential initial false negatives.

2. US ∩ Mx, a believe the negative protocol with US as a triage for Mx. The protocol will have better specificity than Mx alone and in practice would greatly reduce the number of mammograms used in a screening program.

The MLEs for the sensitivity and specificity of these diagnostic protocols under both a standard analysis (without baseline data) and an augmented analysis (with the inclusion of baseline data) are shown in Table 2. Here, we are interested in assessing the gain in precision in estimating the sensitivity and specificity of the new protocols when we include the baseline data in the analysis. We note that, for the believe the positive protocol, the estimates of both sensitivity and specificity are similar with and without the inclusion of the baseline data (around 96% sensitivity and 80% specificity). Also, the precision of the estimator of sensitivity does not change noticeably when we include the baseline data. However, the standard error for the estimate of the specificity decreases by 30%. For the believe the negative protocol, the estimates given by the two analyses are also similar. There is a decrease of 10% in the standard error of the sensitivity estimate and 25% for the specificity estimate when the baseline data are included in the analysis.

For both protocols, the estimates of the logarithm of the ratio of positive and negative likelihood ratios and their associated standard errors for the standard and augmented analysis are also given in Table 2. For the believe the positive protocol, the MLE for the logarithm of the ratio of positive likelihood ratios, $\phi^+_{\text{Mx}\cup\text{US}}$, is negative for both the standard and augmented analyses, although it is more extreme for the augmented analysis (i.e., -0.290 versus -0.359). The standard error of the estimate of $\phi^+_{\text{Mx}\cup\text{US}}$ decreases by around 17% when the baseline data are used in estimation. The estimate of the log-ratio of the negative likelihood ratios for the believe the positive protocol, $\phi^-_{\text{Mx}\cup\text{US}}$, is identical for the standard and augmented analyses (i.e., -1.655). The corresponding standard errors are very similar, with a slight decrease when the baseline data are used (about 4%).

When ultrasound is included in a believe the negative protocol, the MLE of the log-ratio of the negative likelihood ratios , $\phi^-_{\text{US}\cap\text{Mx}}$, given by the augmented analysis is smaller than the one given by the standard analysis (0.347 compared to 0.363). The standard error of the estimate decreases by 17% when the baseline data are used. The estimates of the log-ratio of the positive likelihood ratios, $\phi^+_{\text{US}\cap\text{Mx}}$, are identical (i.e. 1.081) for the standard and augmented analyses. However, the corresponding asymptotic standard error is slightly larger for the augmented analysis than for the standard analysis. To investigate this counterintuitive result, we conducted a simulation study in which the true parameter values and sample sizes are assumed to equal the relevant proportions from the three 2×2 tables in Table 1. The results, given in Table 3, indicate that the asymptotic standard error from the standard analysis underestimates the standard deviation corresponding to $\hat{\phi}^+_{\text{US}\cap\text{Mx}}$, (0.303 versus 0.342). Recall that the expressions for $\hat{\phi}^+_{\text{US}\cap\text{Mx}}$, (also $\hat{\phi}^-_{\text{Mx}\cup\text{US}}$) and its asymptotic standard deviation do not depend on the baseline data. We note a similar situation for the standard error of the estimate of $\phi^-_{\text{Mx}\cup\text{US}}$ from both the augmented and standard analysis

	Sensitivity	Specicity	ϕ^+	ϕ^-
	Believe the positive protocol (Mx ∪ US)			
Standard Analysis	0.958 (0.013)	0.790 (0.027)	-0.290 (0.123)	-1.655 (0.289)
Augmented Analysis	0.955 (0.013)	0.808 (0.019)	-0.359 (0.102)	-1.655 (0.279)
	Believe the negative protocol (US ∩ Mx)			
Standard Analysis	0.617 (0.031)	0.966 (0.012)	1.081 (0.303)	0.363 (0.083)
Augmented Analysis	0.602 (0.028)	0.971 (0.009)	1.081 (0.331)	0.347 (0.069)

Table 2: Estimated sensitivity, specificity, and log-ratio of likelihood ratios (ϕ^+ and ϕ^-) and (standard errors) for mammography (Mx) versus believe the positive (Mx ∪ US) and believe the negative (US ∩ Mx) protocols.

	Sensitivity			Specificity			ϕ^+			ϕ^-		
	Est	ESE	ASE	Est	ESE	ASE	Est	ESE	ASE	Est	ESE	ASE
				Believe the positive protocol (Mx U US)								
Standard Analysis	0.958	0.013	0.013	0.790	0.027	0.027	-0.298	0.127	0.123	-1.702	0.317	0.289
Augmented Analysis	0.955	0.014	0.013	0.808	0.019	0.019	-0.353	0.103	0.102	-1.701	0.317	0.279
				Believe the negative protocol (US ∩ Mx)								
Standard Analysis	0.617	0.031	0.031	0.966	0.012	0.012	1.132	0.342	0.303	0.366	0.084	0.083
Augmented Analysis	0.602	0.028	0.028	0.971	0.009	0.010	1.132	0.344	0.331	0.348	0.069	0.069

Table 3: Estimated sensitivity, specificity and log-ratio of likelihood ratios (ϕ^+ and ϕ^-), empirical standard errors (ESE) and asymptotic standard errors (ASE) for mammography (Mx) versus believe the positive (Mx U US) and believe the negative (US ∩ Mx) protocols from 50,000 simulations based on parameter values derived from the breast cancer example.

which underestimates the corresponding standard deviation (0.289 versus 0.317 for the standard analysis). The simulation study also suggests that the other asymptotic standard errors reported in Table 2 are good approximations to their corresponding standard deviations.

Simulation Study

In the previous section, for the breast cancer screening example, we found that including baseline information generally improved the precision of the estimators of sensitivity and specificity of the new protocols, and of the logarithm of the ratio of positive likelihood ratios (for believe the positive) and negative likelihood ratios (for believe the negative). In order to investigate the magnitude of the potential precision gains across a variety of baseline and assessment study sizes and diagnostic accuracies, we conducted a factorial simulation study.

We considered four study sizes for the baseline data m_0, m_1=100, 250, 1000, 5000 and two sizes for the paired assessment study n_0, n_1=100, 250. Diagnostic accuracies of 70%, 80%, 90% were considered for each of $sens_E$, $spec_E$, $sens_N$, and $spec_N$. In addition, the level of agreement between the two tests, represented through $\pi_{11}^1 = \Pr[E=1, N=1|D=1]$ and, $\pi_{00}^0 = \Pr[E=0, N=0|D=0]$ were set at 10%, 50%, 90% of the way between their minimum and maximum values given by (25) and (26):

$$\max(sens_E + sens_N - 1,\ 0) \le Pr(E = 1,\ N = 1 \mid D = 1) \le \min(sens_E, sens_N) \quad (25)$$

$$\max(spec_E + spec_N - 1,\ 0) \le Pr(E = 0,\ N = 0 \mid D = 0) \le \min(spec_E, spec_N) \quad (26)$$

This leads to a total of 46,656 scenarios. In each scenario the ratio of the asymptotic standard deviation of the estimators of the sensitivity, specificity, ϕ^+ and ϕ^- for an augmented (with baseline data) versus standard (no baseline data) analysis were calculated. Results showing the ratios of asymptotic standard deviations for a believe the positive protocol are presented in Figure 1 for varying baseline sample sizes m_1 and m_0. Results are similar for a believe the negative protocol with increased precision seen for ϕ^- rather than ϕ^+. Over the scenarios considered, the median reductions in asymptotic standard deviation for the sensitivity, specificity, and ϕ^+ were 11%, 19%, and 19%, respectively.

The greatest precision improvements are found with larger baseline study sizes with the number of diseased m_1 impacting the precision of $sens_{E\cup N}$, the number of non-diseased impacting the precision of $spec_{E\cup N}$, and both jointly impacting the precision of $\phi^+_{E\cup N}$. The gains in precision in the estimation of $sens_{E\cup N}$ also tend to be higher when the number of diseased subjects in the assessment study n_1 is small, $sens_E$ is high, $sens_N$ is low, and the level of agreement between $sens_E$ and $sens_N$ is high. For $spec_{E\cup N}$ the greatest gains in precision are found when n_0 is small, $spec_E$ is low, $spec_N$ is high, and their agreement is high. Conversely, the gains in precision in the estimation of $\phi^+_{E\cup N}$ are greatest when $spec_E$ is high, $spec_N$ is low, and their agreement is low.

Discussion

In this paper, we provided methods for incorporating available baseline information into the estimation of the sensitivity and specificity of a new believe the positive or believe the negative protocol, and of the logarithm of the ratios of positive and negative likelihood ratios of the new protocol vs. the existing test. Comparison between the likelihood ratios of the new protocol and the existing test is equivalent to a comparison between the corresponding predictive values when applied to the same population. Asymptotic variances were derived for all proposed statistics.

Throughout the paper, we assumed that baseline data for the augmented analysis were available in the form of a 2×2 table. Alternatively, baseline data might only be available in the form of estimated sensitivity and specificity and their standard errors or confidence intervals. Provided the number of subjects is given, it is straightforward to transform a test's sensitivity and standard error to the data m_e^d, e=0,1 and d=0,1 using formulas for the MLE and variance of a proportion from a binomial distribution. Additionally, the methods presented here could be expanded to include a second set of baseline data perhaps from a preliminary evaluation of the new test.

Before conducting an augmented analysis we propose using a likelihood ratio test of the equivalence of the diagnostic properties of the existing test across the baseline and assessment study data. If this test is rejected it suggests that the existing test is behaving differently between subjects in the baseline and assessment studies. This could be due to inherent differences in the populations (for example age, disease strain, disease severity) or differences in the application of the test. In this case it is unwise to incorporate the baseline data in the analysis and a standard analysis should be used by substituting m_1=m_0=0 into the formulas given in the Methods section.

Incorporating baseline data increases the precision of the estimators of the sensitivity and specificity of the protocol and of the log-ratio of negative (positive) likelihood ratios of the believe the positive (negative) protocol versus the existing test. We do not have an intuitive explanation as to why the MLE and asymptotic variance of $\phi^-_{E\cup N}$ and $\phi^+_{E\cup N}$ do not depend on the baseline data. However, in a believe the positive protocol, where we get higher sensitivity due to an increased number of positive results, we may be more interested in comparing the positive predictive values of $E \cup N$ and the existing test through $\phi^+_{E\cup N}$ whose estimation and precision does benefit from the inclusion of baseline data.

In terms of planning an assessment study, there are two key ways in which incorporating available information is beneficial. First, for

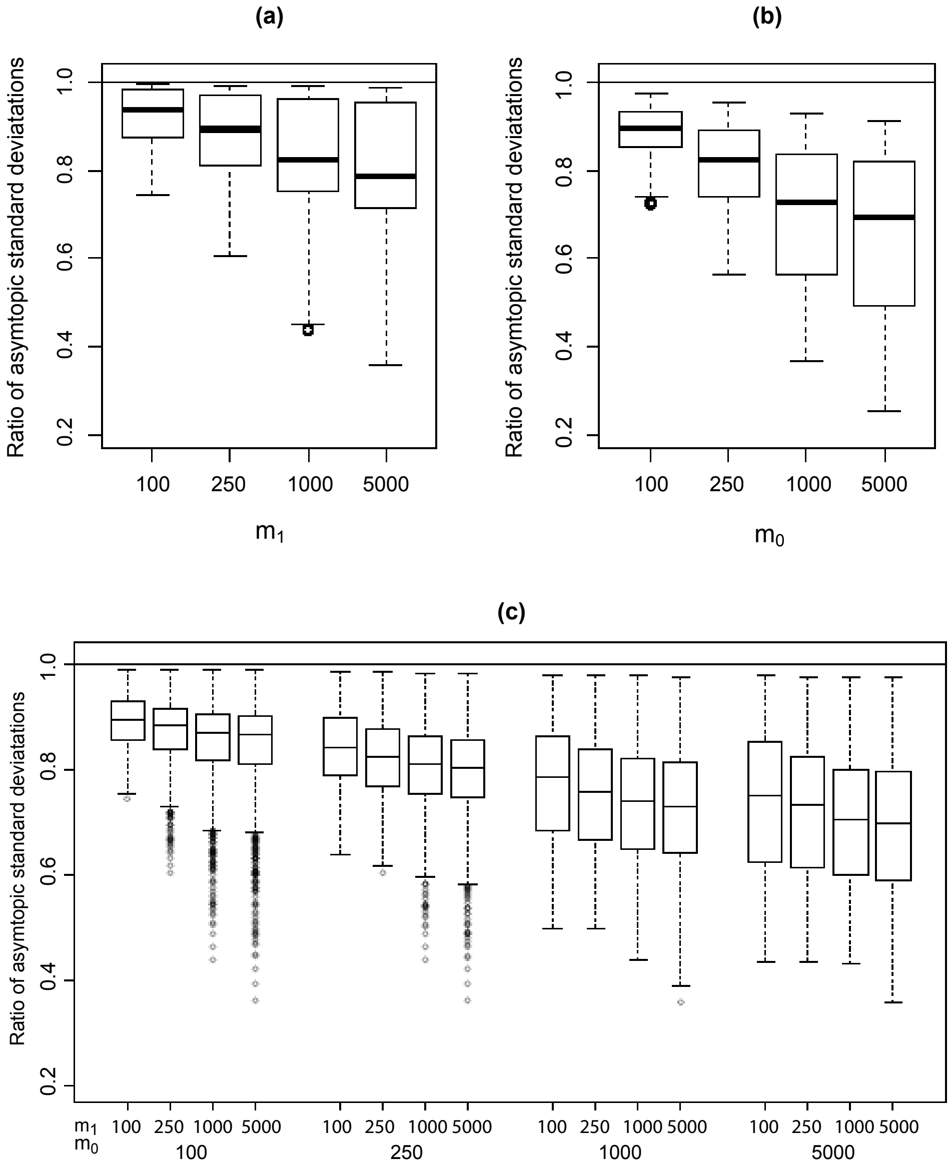

Figure 1: Ratio of the asymptotic standard deviation for an augmented analysis versus standard analysis for estimators of (a) sensitivity of *EUN*, (b) specificity of *EUN*, and (c) ϕ^{+}_{EUN} for a believe the positive protocol across varying baseline sample sizes m_1 and m_0, assessment study sizes, and sensitivities and specificities of the new and existing tests.

a fixed assessment study size, using baseline data will improve the precision of the estimators of important quantities for the combined protocol. Second, for a fixed desired precision, the use of baseline data may substantially decrease the required sample sizes for the assessment study. Initial estimates of the properties of the new and existing test along with the expressions for asymptotic variances given in the Methods section can be used for planning purposes.

If the assessment study is run using a paired design the properties of all four combinations of protocol (believe the positive or believe the negative) and plan (add-on or triage) can be estimated. If the protocol is fixed a priori a sequential assessment study can be run with the existing test first applied to all subjects. This study mimics the use of an add-on plan and gives all the necessary information to compare the existing test with the new protocol, for both augmented and standard analysis. The results apply to either plan. However, in a sequential study mimicking a triage plan only the subjects testing positive with the new test are further tested with the existing test. Therefore, the sensitivity and specificity of the existing test cannot be estimated from the assessment study and therefore cannot be compared to the baseline results. An augmented analysis could still be conducted using methods similar to those developed in the Methods section.

In the example with breast cancer detection, neither the existing test nor the believe the positive protocol outperformed the other in terms of both positive and negative predictive values. In such cases

Macaskill *et al.* [7] propose choosing between the existing test and the protocol based on the expected numbers of additional false positive and true positive results identified by the believe the positive protocol. As this protocol involves testing with the new test the subjects who tested negative for the existing test, the believe the positive protocol will generate some additional false positives and true positives when compared to the results of the existing test only. For a believe the negative protocol one would compare the expected numbers of additional false negative and true negative results identified by the protocol. We do not further investigate this trade-off here but it may warrant future attention.

Supplementary Materials

Appendix A gives the derivation of a likelihood ratio test for testing the equivalence in the diagnostic properties (sensitivity and specificity) of the existing test across the baseline data and assessment study. Appendix B gives the derivation of the asymptotic variances of the estimators of the sensitivity, specificity and log-ratios of the positive and negative likelihood ratios for a believe the positive protocol.

Acknowledgments

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

1. Peña BMG, Taylor GA, Fishman SJ, Mandl KD (2002) Effect of an imaging protocol on clinical outcomes among pediatric patients with appendicitis. Pediatrics 110: 1088-1093.
2. Kotaniemi-Talonen L, Nieminen P, Anttila A, Hakama M (2005) Routine cervical screening with primary HPV testing and cytology triage protocol in a randomised setting. British Journal of Cancer 93: 862-867.
3. Gyselaers WJA, Roets ERA, Van Holsbeke CDYJ, Vereecken AJ, et al. (2006) Sequential triage in first trimester may enhance advanced ultrasound scanning in population screening for trisomy 21. Ultrasound in Obstetrics and Gynecology 27: 622-627.
4. Vishnuvajjala RL (2006) Statistical Review Quality Assessment For Diagnostic PMA Submissions. JSM Proceedings, Section on Statistics in Epidemiology. Alexandria, VA: American Statistical Association. 2640-2644.
5. Zhou XH, Obuchowski NA, McClish DK (2011) Statistical Methods in Diagnostic Medicine. John Wiley and Sons.
6. Marshall RJ (1989) The predictive value of simple rules for combining two diagnostic tests. Biometrics 45: 1213-1222.
7. Macaskill P, Walter SD, Irwig L, Franco EL (2002) Assessing the gain in diagnostic performance when combining two diagnostic tests. Statistics in Medicine 21: 2527-2546.
8. Hayen A, Macaskill P, Irwig L, Bossuyt P (2010) Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. Journal of Clinical Epidemiology 63: 883-891.
9. Nofuentes JAR, del Castillo JDL (2007) Comparison of the likelihood ratios of two binary diagnostic tests in paired designs. Statistics in Medicine 26: 4179-4201.
10. Muwonge R, Walter SD, Wesley RS, Basu P, Shastri SS, et al. (2007) Assessing the gain in diagnostic performance when two visual inspection methods are combined for cervical cancer prevention. Journal of Medical Screening 14: 144-150.
11. Bossuyt PM, Irwig L, Craig J, Glasziou P (2006) Diagnosis: Comparative accuracy: assessing new tests against existing diagnostic pathways. BMJ: British Medical Journal 332: 1089-1092.
12. Ahmed AE, McClish DK, Schubert CM (2011) Accuracy and cost comparison in medical testing using sequential testing strategies. Statistics in Medicine 30: 3416-3430.
13. Ahmed AE, Schubert CM, McClish DK (2013) Reducing cost in sequential testing: a limit of indifference approach. Statistics in Medicine 32: 2715-2727.
14. Danila O, Steiner SH, MacKay RJ (2010) Assessment of a binary measurement system in current use. Journal of Quality Technology 42: 152-164.
15. Danila O, Steiner SH, MacKay RJ (2012) Assessing a binary measurement system with varying misclassication rates using a latent class random effects model. Journal of Quality Technology 44: 179-191.
16. Houssami N, Irwig L, Simpson JM, McKessar M, Blome S, Noakes J (2003) Sydney breast imaging accuracy study: comparative sensitivity and specificity of mammography and sonography in young women with symptoms. American Journal of Roentgenology 180: 935-940.
17. Kavanagh AM, Giles G, Mitchell H, Cawson JN (2000) The sensitivity, specificity, and positive predictive value of screening mammography and symptomatic status. Journal of Medical Screening 7: 105-110.