Austin Biometrics and Biostatistics

Open Access

Austin Publishing Group

Special Article - Biostatistics Theory and Methods

Believe the Extreme (BE) Strategy at the Optimal Point: What Strategy will it become?

Ahmed AE[1], McClish DK[2]*, Schubert CM[3] and AL-Jahdali HH[4]

[1]Department of Epidemiology and Biostatistics, King Saud bin Abdulaziz University for Health Sciences, Saudi Arabia

[2]Department of Biostatistics, Virginia Commonwealth University, USA

[3]Department of Mathematics and Statistics, Air Force Institute of Technology, USA

[4]Department of Medicine, Pulmonary Division-ICU, King Saud bin Abdulaziz University for Health Sciences, Saudi Arabia

***Corresponding author:** McClish DK, Department of Biostatistics, School of Medicine, Virginia Commonwealth University, Virginia, USA

Received: June 01, 2015; **Accepted:** June 11, 2015; **Published:** June 29, 2015

Abstract

The choice of what tests to sequence is essential in making a clinical decision. A variety of sequential techniques have been proposed to combine tests to increase the overall accuracy, including Believe the Positive (BP), Believe the Negative (BN), and the relatively new Believe the Extreme (BE). For a two test sequence, the BP strategy administers Test 2 only if the results on Test 1 are not positive. Similarly, the BN strategy administers Tests 2 only if the results on Test 1 are not negative. For both of these strategies (BP and BN), two thresholds are required. In the BE strategy, only those subjects who tested neither positive nor negative for disease with Test 1 are administered Test 2. Thus there are 3 thresholds for a two test BE strategy: 2 for the initial test, and 1 for the second test. The BE strategy can at times approximate the BP and BN strategies if the upper threshold on the first test is estimated very high or low. This paper explores the BE strategy while varying parameters associated with the features of each test to determine when the BE strategy behaves as a BP or BN strategy, as opposed to requiting all three thresholds. Two practical examples are presented: sleep apnea data and pancreatic cancer. The sleep apnea study shows that the BE strategy might actually function as a BN strategy. The cancer study shows how BE can display better accuracy and lower cost than either the BN or BP strategies.

Keywords: Sequential testing; Believe the extreme; Believe the positive; Believe the negative; Obstructive sleep apnea; Pancreatic cancer

Abbreviations

BP: Believe the Positive; BN: Believe the Negative; BE: Believe the Extreme; PSA: Prostate Specific Antigen; MROC curve: Maximum Receiver Operating Characteristic; FPR: False Positive Rate; U: Upper; L: Lower; N: Non-diseased; D: Diseased; CDFs: Cumulative Distribution Functions; P(D): Prevalence; GYI: Generalized Youden Index; OOP: Optimal Operating Point; b: Standard Deviation Ratio; ρ: Correlations; AUC: Area Under the Curve; OSA: Obstructive sleep apnea; AHI: Apnea-Hypopnea Index; ESS: Epworth Sleepiness Scale; BMI: Body Mass Index; CA125: Cancer Antigen 125; CA19-9: Carbohydrate Antigen 19-9

Introduction

A number of techniques have been used to improve the overall accuracy of the diagnostic process. Sequential testing techniques combine multiple tests sequentially in order to classify subjects into one of two groups (diseased or non-diseased). The use of combinations of logic rules is a popular technique for combining a sequence of tests [1-3]. Two such logic rules, Believe the Negative (BN) and Believe the Positive (BP) are often explored in the literature and currently are the most popular techniques for combining sequential tests [4-7]. For a two test sequence, the BP strategy administers Test 2 to subjects only if the results on Test 1 are not positive. Similarly, the BN strategy for a two test sequence administers Tests 2 to subjects only if the results on Test 1 are not negative. For both of these strategies (BP and BN), two thresholds are required: one for Test 1 and one for Test 2. A relatively new sequential method, which we call Believe the Extreme (BE) is rarely mentioned in the literature [2,3,5]. In the BE strategy with two tests there are 2 thresholds to classify patients as positive, negative, or uncertain for disease on the first test. Patients who test neither positive nor negative for disease in Test 1 are administered Test 2, where a single threshold determines positivity of the test [5]. Etzioni et al. appear to be the first to formalize the statistical evaluation of the BE strategy in the context of prostate cancer (PSA and percent-free PSA) although the strategy was not named by the researchers. Previous studies found that the BE strategy was the most consistently accurate and least costly choice (the cost of testing defined as the number of subjects who need more than one test to diagnose disease) when compared to the BP and BN strategies [5,8]. The BE strategy has also been shown to resolve to a BP or BN strategy as a special case [5]. The BE strategy has also been shown to resolve to a BP or BN strategy as a special case [5].

This paper examines the BE strategy to determine when the BE strategy reduces to the BN or BP strategy for which only a single threshold for the initial test is needed and when two thresholds are required for the initial test. That is, are there scenarios or contexts for which the BE strategy stands on its own, or should researchers only consider the BP or BN strategy. In this investigation, the BE strategy is assessed at the optimal point, considering test characteristics such as ratios of the standard deviations of diseased and non-diseased populations, area under the curve, correlation between the two tests of diseased and non-diseased populations, and prevalence of disease.

Method

The use of the BE strategy is associated with three thresholds (two

Citation: Ahmed AE, McClish DK, Schubert CM and AL-Jahdali HH. Believe the Extreme (BE) Strategy at the Optimal Point: What Strategy will it become?. Austin Biom and Biostat. 2015;2(3): 1022.

thresholds for the first test and one threshold for the second test). The first test (Test 1) is measured on all subjects. When the result of Test 1 is in a grey zone (where subjects cannot be classified as negative or positive), Test 2 is administered to determine their diagnostic status. Specifically, the BE strategy will classify a subject as having disease if the result of Test 1 exceed an Upper threshold (U) or if the result for Test 1 is neither positive nor negative (grey zone) and the second test is positive for disease. The BE strategy will classify a subject as non-diseased if the result of Test 1 is less than a Lower (L) threshold or if the result for Test 1 is neither positive or negative (grey zone) and the second test is negative for disease. This procedure may produce more than one sensitivity value corresponding to a fixed specificity, a result of choosing different thresholds for the BE strategy. The Maximum Receiver Operating Characteristic (MROC) curve has been used to summarize the accuracy results for the BE testing strategy as it depicts the best (maximum) sensitivity for a fixed False Positive Rate (FPR=1-specificity) [5,8].

Computing sensitivity and specificity for the BE strategy

Let X_{1D} and X_{2D} represent the continuous test results of the diseased (D) population for Tests 1 and 2 respectively. Let X_{1N} and X_{2N} represent the test results of non-diseased (N) population for Tests 1 and 2 respectively. Let θ_{1U} and θ_{1L} represent the two thresholds associated with Test 1 where $\theta_{1U} > \theta_{1L}$, and θ_2 is the threshold associated with Test 2. Let F_{1D}, F_{2D}, F_{1N} and F_{2N} represent the Cumulative Distribution Functions (CDFs) of test results for those with (D) and without (N) disease for the first (1) and second (2) test. Finally, let $F_{1D,2D}$ and $F_{1N,2N}$ represent the joint CDFs of test results for those with (D) and without (N) disease between Tests 1 and 2. The BE strategy uses combinations of "AND" and "OR" statements to define overall disease positive or negative test results. The key rules of testing for this strategy are the following:

Positive result if $X_1 > \theta_{1U}$ or $X_2 > \theta_2$ and $\theta_{1L} < X_1 < \theta_{1U}$

Negative result if $X_1 < \theta_{1L}$ or $X_2 < \theta_2$ and $\theta_{1L} < X_1 < \theta_{1U}$

The formula for FPR and Sensitivity (Se) of the BE strategy are given by [5].

$$FPR^{BE}(\underset{\sim}{\theta}), = 1 - F_{1N}(\theta_{1L}) + F_{1N,2N}(\theta_{1L}, \theta_2) - F_{1N,2N}(\theta_{1U}, \theta_2)$$

$$Se^{BE}(\underset{\sim}{\theta}) = 1 - F_{1D}(\theta_{1L}) + F_{1D,2D}(\theta_{1L}, \theta_2) - F_{1D,2D}(\theta_{1U}, \theta_2)$$

Computation of MROC, cost and optimal operating point

When considering two continuous tests, the use of the BE strategy is associated with three thresholds and as such, produces a collection of FPR-Se pairs from which the Maximum Receiver Operating Characteristic (MROC) curve may be derived [5]. These FPR-Se pairs may, and often do, contain values for which at a given FPR=t, multiple Se values are observed. Clearly, thresholds that can produce a higher Se for a fixed FPR may be preferred over those that produce lower Se values. Thus, the MROC curve is comprised of the FPR-Se pairs for which She is maximized at a fixed FPR=t. The formula used to calculate the MROC curve in general is:

$$\text{MROC} = \left\{ \left(\left(t, \max_{FPR(\underset{\sim}{\theta}) \le t} (\text{Se}(\underset{\sim}{\theta})) \right) : 0 < t < 1, \underset{\sim}{\theta} \in \mathbb{R}^n \right) \right\}.$$

For each point on the MROC curve for the BE strategy, there is a corresponding set of thresholds in $\mathbb{R}^3$ (θ_{1L} and θ_{1U} for Test 1 and θ_2 for Test 2) that produces maximum sensitivity for an associated fixed FPR.

Also associated with the set of thresholds that define the points on the MROC curve is a cost of testing. This cost is a measure of the number of subjects that must be evaluated by both tests, and therefore is a function of the probability of the second test being used (i.e., the thresholds on the first test that force the subject to proceed to the second test). The more subjects being classified by the second test, the higher the cost associated with conducting the sequence. Thus, the thresholds associated with the points on the MROC curve make the BE strategy less or more expensive depending on the number of patients who receive Test 2.

The formula used to calculate this cost, the cost of conducting the sequence, is

$$C(\theta_{1L},\theta_{1U}) = ((F_{1D}(\theta_{1U}) - F_{1D}(\theta_{1L}) \times P(D)) + ((F_{1N}(\theta_{1U}) - F_{1N}(\theta_{1L}) \times (1 - P(D)))$$

The MROC curve describes the best performance (highest Se value) across every FPR, and notably, across all threshold combinations. It may be prudent, though, to define and work with the point at which classification accuracy is optimized, that is, rather than working with the entire MROC curve that was generated by the testing sequence, describe instead the performance of the testing sequence at its optimal point. We consider the Optimal Operating Point (OOP) which maximizes the Generalized Youden Index (GYI) where the GYI is given by the following formula:

$$GYI(\underset{\sim}{\theta}) = \max_{\underset{\sim}{\theta}} \left(Se(\underset{\sim}{\theta}) - mFPR(\underset{\sim}{\theta}) \right)$$

Where $m=[(1-P(D)/P(D)] \times [(C_{FP}-C_{TN})/(C_{FN}-C_{TP})]$ and the terms C_{FP}, C_{TN}, C_{FN}, C_{TP} refer to the costs of misclassification associated with a False Positive (FP), True Negative (TN), false negative (FN) and True Positive (TP). The term m, is a weighting factor which represents the slope of the MROC curve at the OOP [8-10]. The misclassification costs in m reflect financial or health costs that result from the decisions of the sequence [9-11], not to be confused with the cost of conducting the sequence, $C(\theta_{1L}, \theta_{1U})$.

A maximum GYI may also be computed amongst the sets of thresholds that restrict cost to particular ranges of values, $C(\theta_{1L},\theta_{1U}) < C_0$ [5]. C_0 Would be a cost restriction, which does not allow the user to consider threshold values, or testing performance, for subsets of patients receiving both Test 1 and Test2 whom exceed a particular cost (e.g. $C_0 = 80\%$ means that no more than 80% of patients would undergo Test 2). A cost constraint of 100% means that all of the patients could receive both Test 1 and Test 2.

Simulation

Simulation methods

To be able to study the behavior of the BE strategy, the effects of four different parameters associated with the accuracy and cost were examined. These parameters were the ratio of the standard deviations for the diseased, σ_D and non-diseased populations, σ_N, prevalence of disease P(D), the correlation between tests in the sequence for the non-diseased, ρ_N, and diseased, ρ_D, populations, and the area under the curve (AUC) for each of the two tests when used alone. In this investigation, the values of the AUC considered for Test 1 and Test 2 respectively, were (0.7, 0.7), (0.7, 0.9) and (0.9, 09). These pairs of values assume that the second test was at least as accurate as the first test. In order to see clearly the effect of the ratio of standard

Table 1: Strategy at the optimal operating point for different values of standard deviation ratio (b), correlations(ρ), areas (AUC) and *m**; no cost restrictions.

		$b_1=b_2=0.5$	$b_1=b_2=1$				$b_1=b_2=2$	$b_1=1, b_2=0.5$	$b_1=1, b_2=2$	$b_1=0.5, b_2=1$	$b_1=2, b_2=1$
		ρ_N, ρ_D	ρ_N, ρ_D				ρ_N, ρ_D	ρ_N, ρ_D	ρ_N, ρ_D	ρ_N, ρ_D	ρ_N, ρ_D
AUC	**m**	(0,0) (.3,.7) (.7,.3) (.7,.7)	(0,0)	(.3,.7)	(.7,.3)	(.7,.7)	(0,0) (.3,.7) (.7,.3) (.7,.7)	(0,0) (.3,.7) (.7,.3) (.7,.7)	(0,0) (.3,.7) (.7,.3) (.7,.7)	(0,0) (.3,.7) (.7,.3) (.7,.7)	(0,0) (.3,.7) (.7,.3) (.7,.7)
(.7,.7)											
	0.5	BP	BE	BE	BP	BE	BN	BE	BE	BP	BE
	1.0	BP	BE	BN	BP	BE	BN	BE	BE	BP	BN
	1.5	BP	BE	BN	BN	BE	BN	BE	BE	BP	BN
(.7,.9)											
	0.5	BP	BE	BE	BP	BE	BN	BP	BN	BP	BN
	1.0	BP	BE	BE	BE	BE	BN	BP	BN	BP	BN
	1.5	BP	BE	BN	BN	BE	BN	BP	BN	BP	BN
(.9,.9)											
	0.5	BP	BE	BE	BE	BE	BN	BE	BE	BP	BE
	1.0	BP	BE	BE	BE	BE	BN	BE	BE	BP	BN
	1.5	BP	BE	BE	BE	BE	BN	BE	BE	BP	BN

**m=[(1-P(D))/ P(D)]×[(C_{FP}-C_{TN})/(C_{FN}-C_{TP})].*

deviations, $b= \sigma_N/\sigma_D$, on the BE strategy, we examined three possible values of the ratio of diseased and non-diseased standard deviations: b=0.5, 1 or 2. When b=2, the standard deviation of test results for non-diseased subjects is twice that of the diseased subjects. When b=1 the standard deviation of test results for diseased and non-diseased subjects are equal. When b=0.5 the standard deviation of test results for diseased subjects is twice that of the non-diseased subjects. Four combinations of correlation between the tests for both the diseased and non-diseased populations were considered:

(ρ_D=0, ρ_N=0), (ρ_D=0.3, ρ_N=0.7), (ρ_D=0.7, ρ_N=0.3) and (ρ_D=0.7, ρ_N=0.7).

Finally, these parameter settings were examined when imposing a cost constraint of 80%, that is, the cost of conducting the sequence was restricted to no more than 80% of subjects being diagnosed by the second test. For the cost constraint comparisons, a prevalence of 0.1 was used.

For this simulation, we assumed that the values of the test results for subjects with and without disease *(X_{1D}, $X2_D$, X_{1N}, X_{2N})* followed bivariate normal distributions expressed as follows:

$$(X_{1N}, X_{2N}) \sim BN\left(\begin{pmatrix}\mu_{1N}\\ \mu_{2N}\end{pmatrix}, \begin{pmatrix}\sigma_{1N}^2 & \rho_N\\ \rho_N & \sigma_{2N}^2\end{pmatrix}\right)$$

$$(X_{1D}, X_{2D}) \sim BN\left(\begin{pmatrix}\mu_{1D}\\ \mu_{2D}\end{pmatrix}, \begin{pmatrix}\sigma_{1D}^2 & \rho_D\\ \rho_D & \sigma_{2D}^2\end{pmatrix}\right)$$

Thus, the estimation of the accuracy measures (sensitivity and FPR) was obtained by using normal distribution CDFs for those with and without disease. Together, these are referred to as the binormal model. Without loss of generality, we set the means of the bivariate normal model for the subjects without disease as 0 and the standard deviations to 1, that is: $\mu_{1N}= \mu_{2N}=0$ and $\sigma_{1N}=\sigma_{2N}=1$. Then $\sigma_{1D}=1/b_1$ and $\sigma_{2D}=1/b_2$ Values of μ_{1D}, μ_{2D} can be obtained for assumed values of AUC by the formula

$$\mu_{iD} = \frac{\sqrt{1+b_i^2}}{b_i}\Phi^{-1}(AUC_i).$$

Correlation values were fixed as one of the parameter settings we varied. Since the binormal model assumption was made, we chose to evaluate thresholds over a grid that ranged between

$$[min(\mu_N-3\sigma_N, \mu_D-3\sigma_D),\ max(\mu_N+3\sigma_N, \mu_D+3\sigma_D)].$$

Because test thresholds had to include values from the distribution of both those with and without disease, this range of test thresholds used the minimum of lower limits of the diseased and non-diseased distributions, and the maximum of the upper limits of diseased and non-diseased distributions. Values outside this range, e.g. when

$\theta_{1L} < \mu_{1N} - 2.56\sigma_N$ *or* $\theta_{1U} > \mu_{1D} + 2.56\sigma_D$, demonstrated less than 0.5% of observations fell in these extremes. When $\theta_{1L} < \mu_{1N} - 2.56\sigma_N$ occurs, the threshold θ_{1L} is so low that the strategy behaves as a BP strategy. Similarly, when $\theta_{1U} > \mu_{1D} + 2.56\sigma_D$ occurs, the threshold θ_{1U} is so large that the strategy behaves as a BN strategy.

Simulation results

The following tables show how the actual strategy at the optimal point can vary according to the AUC's, standard deviations, correlations, or cost restrictions on the cost of conducting the sequence. Table 1 shows the resultant strategy at the optimal point for varied values of AUC, correlation and ratios of the standard deviations between tests. The ratio of standard deviations has the most important effects. When $b_1=b_2=0.5$, the strategy always resolves to a BP strategy, however, when $b_1=b_2=2.0$ it resolves to a BN strategy. This is true regardless of the values of *m* or the correlation. When $b_1=b_2=1$ the situation is different. Usually the BE strategy holds, with two finite thresholds for the first test. This is true when correlation is (0, 0) or (0.7, 0.7). However, when correlations are (0.3, 0.7) or (0.7, 0.3) and $AUC_1=0.7$ either the BP or BN strategy may apply.

When only one of the standard deviation ratios is 1, but the other is either 0.5 or 2.0, the results do not vary by correlation but do depend

Table 2: Strategy at the optimal operating point for different values of standard deviation ratio (b), correlations(ρ), areas (AUC) and *m**, when prevalence is 0.1 and cost is restricted to be less than or equal to 0.8.

		$b_1=b_2=0.5$	$b_1=b_2=1$				$b_1=b_2=2$	$b_1=1, b_2=0.5$		$b_1=1, b_2=2$	$b_1=0.5, b_2=1$		$b_1=2, b_2=1$
		ρ_N, ρ_D	ρ_N, ρ_D				ρ_N, ρ_D	ρ_N, ρ_D		ρ_N, ρ_D	ρ_N, ρ_D		ρ_N, ρ_D
AUC	**m**	(0,0) (.3,.7) (.7,.3) (.7,.7)	(0,0)	(.3,.7)	(.7,.3)	(.7,.7)	(0,0) (.3,.7) (.7,.3) (.7,.7)	(0,0)	(.3,.7) (.7,.3) (.7,.7)	(0,0) (.3,.7) (.7,.3) (.7,.7)	(0,0) (.3,.7) (.7,.3)	(.7,.7)	(0,0) (.3,.7) (.7,.3) (.7,.7)
(.7,.7)													
	0.5	BP	BE	BE	BE	BP	BN	BE	BE	BE	BP	BP	BE
	1.0	BE	BE	BE	BN	BP	BN	BE	BE	BE	BE	BP	BN
	1.5	BE	BE	BE	BN	BN	BN	BE	BE	BE	BE	BE	BN
(.7,.9)													
	0.5	BP	BE	BE	BE	BE	BN	BE	BE	BE	BP	BP	BN
	1.0	BE	BE	BE	BN	BN	BN	BE	BN	BN	BE	BP	BN
	1.5	BE	BE	BE	BN	BN	BN	BE	BN	BN	BE	BE	BN
(.9,.9)													
	0.5	BE	BE	BE	BE	BE	BN	BE	BE	BE	BE	BE	BE
	1.0	BE	BE	BE	BE	BE	BN	BE	BE	BE	BE	BE	BN
	1.5	BE	BE	BE	BE	BE	BN	BE	BE	BE	BE	BE	BN

*$m=[(1-P(D))/P(D)]\times[(C_{FP}-C_{TN})/(C_{FN}-C_{TP})]$.

on the AUC and *m* (Table 1). When $b_1=1$ and the AUC's are the same, either (0.7, 0.7) or (0.9, 0.9), the BE strategy is retained. When AUC= (0.7, 0.9) the strategy is BP when $b_2=0.5$ and BN when $b_2=2.0$. When $b_2=1$ and $b_1=0.5$, or 2.0, the strategies are mostly not BE, except when $m=0.5$ and $b_1=2$. Therefore, the additional threshold that is required by the BE strategy may not be necessary across all cases, as a BP or a BN strategy may be preferred over a BE strategy at the optimal point.

Since results may differ when considering the cost of conducting the sequence, analysis was rerun assessing the optimal point when cost, as computed in Section 2.2, was restricted to be less than or equal to 0.8. Table 2 summarizes these results. For $b_1=b_2=2$, the strategy at the optimal point still functions as a BN strategy. But now when $b_1=b_2=0.5$ all 3 thresholds of the BE strategy were often needed for the optimal point. This was true for all values of *m* when AUC= (0.9, 0.9) and when $m \geq 1$ for the other AUC values. Otherwise, the BP strategy applied.

When cost of conducting the sequence is restricted and only one of the SD ratios (*b*) is 1 the results differ (Table 2). For $b_1=1$ and $b_2=0.5$ or 2 and AUC= (0.7, 0.7) and (0.9, 0.9) the BE strategy is preserved. But when AUC= (0.7, 0.9), and $m \geq 1$ the BN strategy applies. For the case when $b_2=1$ and $b_1=2$, the strategy at the optimal point was sometimes BE and often BN. However, when $b_1=0.5$, the BE strategy holds for all correlations and *m* when AUC= (0.9, 0.9) and for other AUC values when $m=1.5$. Otherwise, BP often applied.

In summary, the BE strategy appears useful for a large number of scenarios, especially under additional constraints of cost of conducting the sequence. For particular scenarios, the upper threshold on the first test is either so high that the sequence functions like a BN strategy or the lower threshold is so low that the sequence functions like a BP strategy. However, the flexibility of fitting a BE strategy incorporates the potential for either of the other strategies (BP or BN).

Applications

Obstructive Sleep Apnea (OSA) during sleep is a fairly common medical problem that, if ignored, may threaten an individual's life [12]. While the prevalence of OSA is reported to be low in American patients (perhaps 1 in 15 adults has OSA of moderate or worse severity) [13,14], OSA is highly prevalent in Saudi Arabia. It was estimated that 40% of the Saudi Arabian people have an interruption of breathing during sleep and do not get enough good sleep [15].

In a study conducted at King Abdulaziz Medical City-Riyadh (KAMC-R), a patient's age, neck size/cm, Body Mass Index (BMI), and daytime sleepiness, as measured by Epworth Sleepiness Scale (ESS) [16] were used to diagnose OSA. (The Arabic ESS version is a reliable and valid scale in screening patients for OSA risk among Arabic-speaking nations) [17,18]. Based on the Apnea-Hypopnea Index (AHI), 869 patients were classified into two groups: 364 (42%) with OSA (AHI≥15) and 505 (58%) with non-OSA (AHI<15) [13,14]. Table 3 has means and standard deviations for those with and without OSA. Patients with OSA had a larger neck size, an older age, a higher BMI, and a higher ESS score as compared to patients without OSA (p-values < 0.05). These measurements were at most modestly correlated with each other (Table 4).

We considered the optimal point for four pairs of tests 1) age and neck size; 2) ESS and age; 3) ESS and BMI; and 4) ESS and neck size. Table 4 lists the optimal point for each of the 4 sequences, along with the GYI, sensitivity, specificity, and cost. Based on the GYI, the best combination, assessed at the OOP with m=1.5, would be age/neck

Table 3: Mean and standard deviation of clinical measurements by sleep apnea status.

Test	Low-risk for OSA (n=505)		High-risk for OSA (n=364)		
	Mean	SD	Mean	SD	p-value
Neck Size	38.70	4.40	41.60	3.50	0. 0001
Age	42.80	17.30	52.30	13.90	0.0001
BMI	35.40	10.90	37.90	8.20	0.0002
ESS	9.00	5.60	10.40	5.80	0.0006

Table 4: Accuracy and cost at optimal operating point for various OSA test combinations.

Test1/Test2	ρ_D / ρ_N	b_1 / b_2	c1	c2	c3	GYI	Sensitivity	FPR	cost
Age/ Neck size	-0.5 / 0.31	1.2 / 1.3	66.8	36.8	40.3	0.155	0.613	0.305	0.623
ESS/Age	0.06 / 0.02	1 / 1.2	27.2*	8.5	49.1	0.088	0.381	0.195	0.757
ESS/Neck size	0.22 / 0.1	1 / 1.3	26.7*	6.5	41.0	0.132	0.454	0.215	0.703
ESS/BMI	0.13/ -0.01	1 / 1.3	26.6*	13.5	33.4	0.046	0.226	0.120	0.245

*The actual upper limit of the EPS survey is 24. A value above this arose from a grid search assuming normality.

Table 5: Optimal operating point[1] (OOP), accuracy measures and cost for CA125 and CA19-9 using 3 sequential strategies.

	BP	BN	BE
OOP	(111.73, 33.51)	(5.37, 30.85)	(6.59,53.38,6.16)
Se/FPR/GYI	0.75/0.13/0.63	0.74/0.12/0.62	0.80/0.13/0.67
Cost	0.95	0.99	0.80

In the original (back-transformed) units: *(θ_1, θ_2) for BP, BN; (θ_{1L}, θ_{1U}, θ_2) for BE.*

size. As compared to age/neck size, the combination ESS/neck size has slightly lower GYI and higher cost. (GYI=0.155 vs. 0.131, cost=0.623 vs. 0.728, respectively). Cost is much less for the combination ESS/BMI than any of the other test combinations examined, but the sensitivity is also considerably lower, providing the lowest GYI. Note that for the strategies with the initial test ESS, the upper threshold for ESS is initially at or above the maximum value of 24 for the survey. This implies that people would not be classified positive on the first test. Thus in these cases, the BE strategy becomes essentially a BN strategy. Note that the ratio of standard deviations for ESS was 1.0 and for the various second tests b_2>1. Thus these results are consistent with our findings in Table 1.

A second example shows the potential usefulness of the BE strategy in comparison with the BP and BN strategies. Data were from a case-control study conducted at the Mayo Clinic in Rochester, MN, in which blood serum was taken from 141 patients (51 controls with pancreatitis but without pancreatic cancer, 90 cases with pancreatic cancer) to study two antigens: CA125, a cancer antigen, and CA19-9, a carbohydrate antigen [19]. The AUC for CA125 was 0.79 and for CA19-9 was 0.88. The data were not normally distributed, so a Box-Cox transformation was needed (The transformation parameters for CA125 and CP19-9 were λ_1 = -0.5 and λ_2 = -0.25 respectively). After transformation, the ratio of standard deviations was b_1=0.94 and b_2=0.62. The optimal points were determined for the 3 strategies BP and BN as well as BE. Results in Table 5 show that the thresholds for CA125 and CA19-9 when using a BP strategy are 111.73 U/ml and 33.51 U/ml. Thus, in this sample when using the BP strategy, only 7 of 141 people (5%) would have a value above 111.73 and be diagnosed using only CA125. For BN, the OOP would be 5.37 U/ml for CA125 and 30.85 U/ml for CA19-9 and only 1 of 141 would be diagnosed based on CA125. Everyone else would be diagnosed based on the CA19-9 result. In contrast, the OOP for the BE strategy would be 6.59 and 53.38 for CA125 and 6.16 for CA19-9. Twenty percent of patients would be diagnosed based on CA125. The cost for BE was also less than that for BP or BN, indicating that the BE strategy was superior to the other two strategies based on both cost and accuracy.

Discussion

In this paper, we studied the accuracy of the BE strategy under different parameter settings to determine when all 3 thresholds were needed for diagnosis at the optimal point. Depending on the values of the two thresholds on the first test, it is possible for the BE strategy to behave similar to either a BN strategy (when the upper threshold on the first test was very high) or a BP strategy (when the lower threshold on the first test was very low). We found that the ratio of the standard deviations of diseased and non-diseased populations, correlation between the two tests of diseased and non-diseased populations, AUC of the individual tests, and the weighting parameter, *m* of the GYI were all important determiners of whether 1 or 2 thresholds were needed for the initial test in the sequence. When no cost restrictions were placed, the BE strategy resolves to a BP strategy for b < 1, while the BE strategy resolves to a BN strategy for b >1. However, the addition of a cost restriction on conducting the sequence did not allow the BE strategy to collapse to either a BP or BN strategy to the extent as when there was no such restriction. When cost restrictions on conducting the sequence exist, the resulting strategy is more complex, and the optimal point of the test sequence evolves from a combination of the parameter settings.

The choice of what tests to sequence is essential in making a clinical decision. This is illustrated by an example of screening for obstructive sleep apnea, where available information included a patient's age, neck size, BMI, and ESS. Our investigation considered which would be a better pair of tests in order to retain high accuracy while maintaining lower costs. The study revealed that sequencing age and neck size to screen for obstructive sleep apnea leads to more accuracy and concurrently reasonable cost compared to the other pairs. The combination of ESS and neck size yielded high accuracy as compared to the other combinations of tests, but proved more costly than the other combinations. These trade-offs will often exist. When ESS was used as Test 1, the BE strategy was found to essentially function as BN.

At times the BE strategy will allow for fewer uses of the second test at the optimal decision point as compared to BP and BN. This was the case with the Wieand data where the BP and BN strategies essentially required all patients to need CA19-9 while only 80% needed CA19-9 results with the BE strategy. If CA19-9 were particularly expensive or burdensome, this would have been an important advantage in the screening for and diagnosis of pancreatic cancer. The BE sequential strategy is a flexible strategy allowing optimal points to include a set of 2 or 3 thresholds. The choice to determine whether or not a BE, BP, or BN strategy would be optimal is complex and depends on a number of features related to the data structure including the values at which the thresholds for Test 1 are considered extreme. However, it is not necessary to decide in advance whether a BP or BN strategy would be preferred, as modeling with a BE strategy will produce the appropriate thresholds to maximize accuracy for a particular application.

Acknowledgement

We would like to thank King Abdulaziz City for Science and Technology for providing funding for the OSA study (Research Protocol #83-84) and King Abdullah International Medical Research Center and King Abdulaziz Medical City-Riyadh for providing scientific institution approval to carry out the study.

References

1. Ruczinski I, Kooperberg C, Leblanc M. Logic regression. J Comp Graph Statist. 2003; 12: 475-511.
2. Baker SG. Identifying combinations of cancer markers for further study as triggers of early intervention. Biometrics. 2000; 56: 1082-1087.
3. Etzioni R, Kooperberg C, Pepe M, Smith R, Gann PH. Combining biomarkers to detect disease with application to prostate cancer. Biostatistics. 2003; 4: 523-538.
4. Marshall RJ. The predictive value of simple rules for combining two diagnostic tests. Biometrics. 1989; 45: 1213-1222.
5. Ahmed AE, McClish DK, Schubert CM. Accuracy and cost comparison in medical testing using sequential testing strategies. Stat Med. 2011; 30: 3416-3430.
6. Thompson ML. Assessing the diagnostic accuracy of a sequence of tests. Biostatistics. 2003; 4: 341-351.
7. Shen C. On the principles of believe the positive and believe the negative for diagnosis using two continuous tests. Journal of Data Science. 2008; 6: 189-205.
8. Ahmed AE, Schubert CM, McClish DK. Reducing cost in sequential testing: a limit of indifference approach. Stat Med. 2013; 32: 2715-2727.
9. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. 3rd edn. Oxford: Oxford University Press. 2005.
10. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic Evaluation in Clinical Trials. Oxford: Oxford University Press. 2007.
11. Willan AR, Briggs AH. Statistical Analysis of Cost-Effectiveness Data. NY: Wiley. 2006.
12. National Heart, Lung, and Blood Institute. Fact Book: Fiscal Year 1993. US Department of Health and Human Services, US. Public Health Service, National Institutes of Health. 1994.
13. Young T, Peppard PE, Gottlieb DJ. Epidemiology of obstructive sleep apnea: a population health perspective. Am J Respir Crit Care Med. 2002; 165: 1217-1239.
14. Young T, Skatrud J, Peppard PE. Risk factors for obstructive sleep apnea in adults. JAMA. 2004; 291: 2013-2016.
15. Bahammam AS, Al-Rajeh MS, Al-Ibrahim FS, Arafah MA, Sharif MM. Prevalence of symptoms and risk of sleep apnea in middle-aged Saudi women in primary care. Saudi Med J. 2009; 30: 1572-1576.
16. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991; 14: 540-545.
17. Ahmed AE, Fatani A, Al-Harbi A, Al-Shimemeri A, Ali YZ, Baharoon S, et al. Validation of the Arabic version of the Epworth sleepiness scale. J Epidemiol Glob Health. 2014; 4: 297-302.
18. Ahmed AE. Validation of Arabic versions of three sleep surveys. Qatar Med J. 2014: 130-136.
19. Wieand S, Gail MH, James BR, James KL. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. Biometrika. 1989; 76: 585-592.

Citation: Ahmed AE, McClish DK, Schubert CM and AL-Jahdali HH. Believe the Extreme (BE) Strategy at the Optimal Point: What Strategy will it become?. Austin Biom and Biostat. 2015;2(3): 1022.