ORCA – Online Research @ Cardiff

Optimal testing policies for diagnosing patients with intermediary probability of disease

Edilson F. Arruda[a], Basílio B. Pereira[a], Clarissa A. Thiers[b], Bernardo R. Tura[b]

[a]*Universidade Federal do Rio de Janeiro, Instituto Alberto Luiz Coimbra de Pós Graduação e Pesquisa de Engenharia, Programa de Engenharia de Produção, Caixa Postal 68507, Rio de Janeiro RJ 21941-972, Brasil.*
[b]*Instituto Nacional de Cardiologia, Rua das Laranjeiras 374, Laranjeiras, Rio de Janeiro, RJ, 22.240-006, Brasil.*

Abstract

This paper proposes a stochastic shortest path approach to find an optimal sequence of tests to confirm or discard a disease, for any prescribed optimality criterion. The idea is to select the best sequence in which to apply a series of available tests, with a view at reaching a diagnosis with minimum expenditure of resources. The proposed approach derives an optimal policy whereby the decision maker is provided with a test strategy for each a priori probability of disease, aiming to reach posterior probabilities that warrant either immediate treatment or a not-ill diagnosis.

Keywords: Healthcare Problems, Stochastic Shortest Path, Diagnosis

*⋆Corresponding author. Tel.: +55 21 3938 8255; fax +55 21 2270 9702.
Email addresses:* `efarruda@po.coppe.ufrj.br` (Edilson F. Arruda *⋆*), `basilio@hucff.ufrj.br` (Basílio B. Pereira), `clarissa@cardiol.br` (Clarissa A. Thiers), `tura@centroin.com.br` (Bernardo R. Tura)

1. Introduction

Physicians are naturally Bayesian (Gill et al., 2005). Upon receiving a patient, they often intuitively process epidemiological data, anamnesis information, as well as the results of physical examination to reach, either implicitly or explicitly, an estimate of the patient's a-priori probability of having a given disease. If that probability is found insufficient to warrant a diagnosis, the physician may find it necessary to prescribe one or more available specialized tests. It is worth pointing out that a probability may be deemed insufficient either because it is too high to discard a disease or too low to confirm it (e.g., Pauker and Kassirer, 1980).

The utility of sequential testing to help clinicians reach a diagnosis has been found in early investigations (e.g, Hall, 1967; Hall et al., 1985; Weintraub et al., 1984), and the Bayesian approach underlying testing and screening remains a current subject of discussion (Hamm and Beasley, 2014). Recently, Lin et al. (2012) proposed an approach to aggregate different test results in order to reach better diagnosis.

Many works in the literature are concerned with finding a good testing strategy for the diagnosis of a given disease. For example, Ferreira et al. (2014) and Bertoldi et al. (2016) devised arbitrary testing strategies for coronary disease and compared these strategies to select one with an adequate performance. Whereas these approaches can guide clinicians in their choice of tests, they fail to examine all possible combinations of tests, given that one has to test each strategy separately. Moreover, the problem is solved in a static fashion and the paths for diagnosis are defined a-priori, regardless of the posterior probability of disease after testing. Such an strategy is often

called *open loop*, as opposed to a *closed loop* strategy, which would consider the current state of the system before reaching a decision on the next test to be taken (e.g., **???**). It is the the latter strategy that we explore in this paper.

This paper proposes a dynamic model for sequential testing, taking a Bayesian approach to sequentially derive the posterior probability of disease, given the a-priori probability and the result of the last test.

Such a model is related to *sequential hypothesis testing problems* (SHTP) (e.g. **???**Cui and Mahajan, 2015), but there are fundamental differences. Firstly, the proposed problem features a limited number of specialized tests and each test can only be applied once; hence, the number of observations is limited. SHTP, in contrast, seek an optimal sequence of sensing actions before reaching a decision on which hypothesis to select; but the actions merely define how the next sample will be acquired and one can take as many samples as needed (**?**).

Another distinguishing feature of the proposed model with respect to SHTP is that the sampling cost is variable, for it depends on the selected specialized test. SHTP, on the other hand, often seek to find a balance between the number of observations and the accuracy of the selected hypothesis, assuming a constant cost of sample acquisition. Indeed, the proposed formulation is more general in that it does not impose any constraint on the structure of the cost function.

The proposed model is similar to that of Cui and Mahajan (2015), who solve a sequential hypothesis testing problem by means of a Markov model with absorption. For the studied problem, the target set can be comprised of

the posterior probabilities of disease which warrant an immediate diagnosis, as proposed by Pauker and Kassirer (1980). These probabilities are either in a low-probability interval, for which the patient is very likely *not-ill* and should only be observed; or in a high-probability interval, for which the patient is very likely *sick* and should commence immediate treatment. Hence, one is now faced with a stochastic sequential decision making problem that is terminated upon reaching the target set, or running out of tests. This problem is known in the literature as *stochastic shortest path* (e.g., Bertsekas, 2012) and can be readily solved by means of a dynamic programming algorithm.

To the best of our knowledge, this is the first study that incorporates a Bayesian approach into a Markov model to determine an optimal testing sequence for clinical diagnosis. This results in a stochastic shortest path model which allows the decision maker to find the optimal testing strategy by means of a dynamic programming algorithm, which implicitly compares all possible testing policies. Moreover, whenever a strategy exists which terminates in the target set, the resulting optimal policy is guaranteed to reach a diagnosis; that may not be the case when the testing strategy is arbitrarily prescribed, as often happens.

A nice by-product of the proposed approach is that it finds the optimal policy with regards to the prescribed optimality criterion, and that policy can be used as a benchmark for the incorporation of new technology in healthcare. For example, if a new test is developed, its cost should be compared to that of the most economical testing policy, to enable decision makers to weigh the potential benefits against the increase in the overall cost up to diagnosis.

4

The proposed approach is illustrated by an evaluation of the available testing strategies for diagnosing coronary artery disease for the Brazilian public health system. We find the optimal testing strategy with respect to total cost and compare it to the optimal strategy with regards to correct diagnosis probability. The results show that, while the most economical strategy is slightly inferior in terms of the probability of correct diagnosis, it produces significant savings with respect to the competing policy.

This paper is organized as follows. Section 2 introduces the problem and develops a mathematical model for it. The proposed formulation is presented in Section 2.1 and the solution procedure is presented in Section 2.2. Section 3 features numerical examples, including an evaluation of testing strategies for coronary artery disease, in the perspective of the Brazilian public health system. Finally, Section 4 concludes the paper.

2. Mathematical model

Suppose that a physician, after carrying out a thorough physical exam and a detailed anamnesis, reaches the conclusion that the patient's a-priori probability of having a given disease is $\rho \in (0, 1)$. Given this probability, two situations arise:

(i) The physician can consider this probability small or large enough to reach a diagnosis, deciding either that the patient is probably not ill and should be observed, or that the patient is likely to be ill and treatment should be initiated;

(ii) Upon considering this probability not sufficient to reach a diagnosis,

5

the physician may recommend that the patient undergo one or more specialized tests.

Whenever she selects the latter alternative, the physician uses the results of the tests to update the previously assigned probability of disease, thus reaching a posterior probability, which may still be perfected by further tests, if necessary. Typically, tests are performed in sequence, up to the moment a diagnosis is reached (e.g., Pauker and Kassirer, 1980); however some strategies may prescribe a maximum number of exams whereby the last exam results in a diagnosis, regardless of the posterior probability of disease resulting from it (e.g., Ferreira et al., 2014; Bertoldi et al., 2016).

For modeling purposes, one may assume that there exist two probability intervals that yield a diagnosis: an interval of small probabilities I_1 and an interval of large probabilities I_2, such that

$$I_1 = [0, b_1), \qquad I_2 = (b_2, 1], \quad 0 < b_1 < b_2 < 1. \tag{1}$$

Whenever the posterior probability of disease falls within the interval I_1, a *not ill* diagnosis is reached; conversely, when the posterior probability of disease belongs to interval I_2, the physician decides that the patient is probably ill and should commence treatment. Finally, whenever none of the previous alternatives occur, further testing is needed.

Assume that the physician can choose from a set of n available tests $T = \{T_1, \ldots, T_n\}$. Suppose also that no single test $T_i \in T$, $i = 1, 2, \ldots, n$, can be applied twice. That means that a second test $T_i \in T$ provides no further evidence on the presence or absence of disease with respect to the first. Assume also that the tests, given the patient's state, are statistically independent. Following the notation in (Pereira and Pereira, 2005), let $P(D_+) = \rho$

6

be the a-priori probability that the patient is ill, also known as *prevalence*. Then, it follows that:

$$\hat{\rho}_+(T_i, \rho) = \frac{P(T_i^+|D_+)\rho}{P(T_i^+|D_+)\rho + P(T_i^+|D_-)(1-\rho)}, \qquad (2)$$

where $\hat{\rho}_+(T_i, \rho)$ is the posterior probability that the patient is *ill* given a positive test T_i and an a-priori probability of disease ρ. In addition, D_+ represents presence of illness, D_- denotes a healthy patient, and T_i^+ represents a positive test T_i.

Finally, we can also write:

$$\hat{\rho}_-(T_i, \rho) = \frac{P(T_i^-|D_+)\rho}{P(T_i^-|D_+)\rho + P(T_i^-|D_-)(1-\rho)}, \qquad (3)$$

where $\hat{\rho}_-(T_i, \rho)$ is the posterior probability that the patient is *ill* given a *negative* test T_i and an a-priori probability of disease ρ, and T_i^- represents a negative test T_i. For a detailed treatment of diagnostic tests in medicine, we refer the interested reader to (Pereira and Pereira, 2005).

2.1. Shortest path formulation

Let $k \geq 0$, $k \in \mathbb{Z}$ be the number of tests performed so far, and let process X_k, $k \geq 0$, be the current probability of disease, ρ, given the results of the k tests performed so far. The state space of process X_k is $S_X = [0, 1]$.

Recalling the intervals in (1), we reach a diagnosis whenever $X_k \in I_1$ or $X_k \in I_2$. Hence, whenever $X_k \in I = [b_1, b_2]$, one should select a new test to be performed. Let Y_k, $k \geq 0$ be a binary vector whose $i-th$ component - $Y_k(i)$ - represents test $T_i \in T$; if $Y_k(i) = 1$, the patient has already undertaken test T_i and it is no longer an option; otherwise, if $Y_k(i) = 0$, this test was

not performed yet and remains as an option. The state space of process Y_k is defined as $S_Y = \{0, 1\}^n$, where n is the number of available tests.

Let $Z_k = (X_k, Y_k)$ be the stochastic process representing the evolution of the diagnosis, and let $S = S_X \times S_Y$ denote its state space. For each state $z = (\rho, y) \in S$, let $A(z) = \{T_i : y(i) = 0\}$ denote the available tests at state z, and let $A = \cup_{z \in S} A(z)$ denote the action space. If $Z_k = z$ and test $a \in A(z)$ is selected, then the process goes to state $Z_{k+1} = w \in S$ with probability

$$
p_{zw}^a = \begin{cases} P(T_a^+|\rho) = \rho P(T_a^+|D_+) + (1-\rho)P(T_a^+|D_-), & \text{if } w = (\hat{\rho}_+(T_i, \rho),\ \bar{y}^a), \\ P(T_a^-|\rho) = \rho P(T_a^-|D_+) + (1-\rho)P(T_a^-|D_-), & \text{if } w = (\hat{\rho}_-(T_i, \rho),\ \bar{y}^a), \\ 0, & \text{otherwise}, \end{cases}
$$

$$(4)$$

where

$$
\bar{y}^a(i) = \begin{cases} y(i), & \text{if } i \neq a, \\ 1, & \text{if } i = a. \end{cases}
$$

Eq. (4) yields that, if test T_a is positive, the next a-priori probability is the posterior probability given T_a^+ and $P(D_+) = \rho$; otherwise, the next a-priori probability of disease is the posterior probability given T_a^- and $P(D_+) = \rho$. With regards to process Y_k, $k \geq 0$, Y_{k+1} equals Y_k except for component a, which is changed from 0 to 1 because test T_a will have been applied at the next step. Note that the $i - th$ component of $\bar{y}_a$ represents test i and is one whenever test i is already taken, and nil otherwise. Since Eq. (4) evaluates the effect of performing test a, no other test will be performed and hence the components of y that do not concern test a will not be changed. In addition, the component that concerns test a is changed to one, since now test a will be effectively taken at the next state.

8

Let $\pi : S \to A$ denote a stationary control policy, which yields a single test T_a to be performed at each state $z \in S$, and let Π be the set of feasible stationary control policies. For the sake of consistency, we add action $a = 0$ to each state z in the set

$$F = \{z = (x, y) \in S : x \in I_1 \text{ or } x \in I_2\}. \tag{5}$$

This action corresponds to selecting no further tests and is the one to be taken in F, the set of states that warrant an immediate diagnosis.

To guide our choice, let us define a positive cost function $c : S \times A \to \mathbb{R}_+$. For each state $z = (x, y) \in S$, $c(z, a)$ represents the cost of taking test a at state z. For $z \notin F$ this cost function may represent, for example, the cost of test a, or any other performance function of interest. For a given state $z \in F$, on the other hand, we only need to define $c(z, 0)$ - no additional test is needed at this state - which represents an end cost, i.e. the cost of reaching diagnosis at state z. Typically, the target states with disease probability closer to 0 or 1 will have lower end costs, for they yield more precise diagnoses.

To guarantee that a diagnosis will be reached whenever possible, we make

$$c(z, a) = M < \infty, \forall a \in A(z), \text{ if } z = (x, y) \notin F \text{ and } y = [1, \dots, 1], \tag{6}$$

where M is a very large penalty, to ensure that a path without diagnosis is always avoided if possible. In addition, we create an artificial cemetery state Δ with $p_{\Delta\Delta}^a = 1$, $\forall a \in A$ and $c^a(\Delta, a) = 0$, $\forall a \in A$. We also make

$$p_{z\Delta} = \begin{cases} 1, & \text{if } z = (x, y) \notin F \text{ and } y = [1, \dots, 1]; \\ 0, & \text{otherwise.} \end{cases}$$

9

This ensures that, whenever the system runs out of tests before reaching a diagnosis, a penalty M is incurred and the system is transitioned into cemetery state Δ. The equation above also implies that the system will transition to the cemetery state whenever the last test is taken, but the resulting posterior probability is not in the diagnosis intervals I_1 or I_2, defined in Eq. (1).

We emphasize that (6) applies to states for which all possible tests have already been performed. Hence, the formulation makes certain that strategies which do not reach a diagnosis are always outperformed by those that do. Consequently, whenever a strategy exists which reaches diagnosis for all intermediate probabilities, we are guaranteed to find an optimal policy with the same property.

For each state $z \in S : x \notin F$, let

$$V^\pi(z) = E\left\{\sum_{k=0}^{\tau} c(Z_k, \pi(Z_k)) | X_0 = z\right\} \tag{7}$$

be the cost-to-go function of state z, where $\tau \triangleq \min\{k \geq 0 : X_k \in (F \cup \Delta)\}$ is the time the system either reaches a diagnosis or runs out of tests under policy π. The objective of the decision maker is to find an optimal policy $\pi^* \in \Pi$ such that

$$V^*(z) \triangleq V^{\pi^*}(z) \leq V^\pi(z), \quad \forall z \in S, \text{ and } \pi \in \Pi. \tag{8}$$

Problem (8) is a classical *stochastic shortest path problem*, and it can be solved by means of a classical value iteration algorithm (e.g., Bertsekas, 2012).

Remark 1. *Constraint* (6) *makes sure that, whenever there is a probability that the bounds for* ill *and/or* not-ill *diagnosis cannot be reached from some*

10

initial state z_0, we will have $V^*(z_0) > M$. This indicates that either the diagnosis thresholds should be updated or a new, more efficient test, should be added to the mix.

2.2. Solution of the shortest path formulations

To solve problem (8), we employ the classical value iteration algorithm (e.g., Bertsekas, 2012), detailed below.

Algorithm 1 (Stochastic shortest path algorithm).

1. $V_0 \leftarrow 0$, *make* $k = 0$, *and choose a tolerance tol.*

2. *For all* $z \in S : z \notin F$, *make*

$$V_{k+1}(z) = \min_{a \in A(z)} \left\{ c(z,a) + \sum_{w \in S} p^a_{zw} V_k(z) \right\},$$

where p^a_{zw} *is determined from Eq. (4) and F is defined in (5).*

3. $\epsilon \leftarrow \|V_{k+1} - V_k\|$, $k \leftarrow k + 1$.

4. *If* $\epsilon > tol$, *return to step 2.*

5. *For all* $z \in S : z \notin F$, $V^*(z) = V_k(z)$

$$\pi^*(z) \leftarrow \arg \min_{a \in A(z)} \left\{ c(z,a) + \sum_{w \in S} p^a_{zw} V_k(z) \right\}.$$

Note that the output of algorithm 1 is a mapping from state to action, which assigns a single action $\pi^*(z)$ to be taken whenever the system visits a given state $z \in S$. The value function obtained in Step 5 corresponds to the optimal expected cost up to diagnosis, starting from $X_0 = z$.

11

Algorithm 1 is a classical dynamic programming routine (e.g., Bertsekas, 2012). To implement it, one needs first to enumerate all possible states of the system. Each state is a vector of $n + 1$ components, containing a probability of disease and n binary values, each representing a distinct test in $T_i \in T$, which will be assigned 1 if T_i is already taken and zero otherwise. Then, for each state, one has to keep a record of all actions available, i.e. tests not already taken, and assign the specific cost for each pair (state, action), denoted by $c(z, a)$. After that, one uses (5) to evaluate p_{zw}^a for all possible $w \in S$ and keeps record of these values. Then, after all parameters are obtained and stored, one can easily implement the algorithm using any standard programming language, such as R, Matlab, C++, etc. The output of the algorithm gives, for each possible intermediate probability of disease and subset of tests taken, the next test that should be taken among those still left.

3. Numerical Examples

In this section, we apply the proposed formulation in two distinct settings. In the first subsection, the approach is employed to compare tests for Coronary artery disease (CAD). The second subsection is concerned with the application of the approach when some of the available tests present random likelihood ratios.

3.1. Testing for Coronary Artery Disease

In this section, we apply the proposed formulation for comparing Coronary artery disease (CAD) tests (Heijenbrok-Kal et al., 2007). The available tests are exercise electrocardiography (Ex-ECG), stress echocardiogra-

phy (ECHO), single-photon emission computed tomography (SPECT), computed tomography coronary angiography (CTA), or stress cardiac magnetic resonance imaging (C-MRI).

Table 1 depicts some characteristics of these tests (**?**). Test threshold and treatment threshold are respectively the values of b_1 and b_2 in Eq. (1), which are specific of the exam. For the example, however, we will employ $I_1 = [0, 0.2)$ and $I_2 = (0.6, 1]$, thus making $b_1 = 0.2$ and $b_2 = 0.6$, which are close to the thresholds of the Ex-ECG exam, used as a benchmark due to its small cost. The values in the last column of Table 1 are the 2016 reimbursement of the Brazilian public health system (SUS) for Ex-ECG, ECHO and SPECT. Unfortunately, CTA and stress C-MRI are not yet incorporated into the collection of procedures financed by SUS. Hence, for the sake of analysis, we estimated CTA and C-MRI costs based on similar currently reimbursed tests, respectively chest CT and rest C-MRI. These costs were adjusted according to the ratios between them and those of the studied exams, as found in the Brazilian private sector. The properties of each exam described in Section 2 are presented in Table 2 (**?**).

Table 1: Characteristics of the Tests.

Test	Test Threshold	Treatment Threshold	SUS Reimbursement (R$)
Ex-ECG	0.22	0.58	30.00
ECHO	0.10	0.72	165.00
CTA	0.06	0.81	328.54
C-MRI	0.16	0.80	361.25
SPECT	0.12	0.80	791.39

Table 2: Properties of the Tests. Source: (**?**)

Test	Sensitivity	Specificity	Positive Likelihood Ratio	Negative Likelihood Ratio
Ex-ECG	0.67	0.71	2.31	0.46
ECHO	0.79	0.87	6.07	0.24
CTA	0.87	0.91	9.60	0.14
C-MRI	0.89	0. 76	3.70	0.14
SPECT	0.87	0.81	4.57	0.16

We solved problem (8) to find the optimal testing strategy for all integer values of intermediate a priori probability of disease (in percentage points), with respect to both the overall cost of diagnosis and the probability of correct diagnosis. To find the optimal strategy with respect to the overall cost, we define

$$c(z, a) = f(a), \ \forall z \in S \setminus F, \ a \in A,$$

in Eq. (7); where $f(a)$ is the cost of exam a, given in the last column of Table 1.

In contrast, to find the optimal policy with respect to the probability of correct diagnosis, i.e. the policy that maximizes the probability of correct diagnosis, we set:

$$c(z, a) = \begin{cases} 0, & \text{if } z \notin F, \\ -(1 - \rho), & \text{if } z \in F \text{ and } x < b_1, \\ -\rho, & \text{if } z \in F \text{ and } x > b_2. \end{cases}$$

In the expression above, the numbers in the right hand side represent the opposite of the posterior probability of correct diagnosis. Note that, if $\rho < b_1$ the diagnosis is *not ill*, but the probability of disease is ρ, therefore, a correct

14

diagnosis will be reached with probability $(1-\rho)$. Correspondingly, if $\rho > b_2$, the diagnosis will be that the patient is ill, but the probability of being ill is actually ρ. The opposites are used because Problem (8) was defined as a minimization problem and our objective is to maximize the objective function.

Table 3 describes the optimal cost policy, comprised of the sequence of tests to be taken for each a-priori probability of disease. Note that this policy prescribes Ex-ECG as the first test for a-priori probabilities in the intervals $[20\%, 31\%]$ and $[40\%, 60\%]$. For the lower probabilities a *not-ill* diagnosis is reached with a negative result, whereas a positive result is inconclusive and demands ECHO as the second test. If the second test is taken, its result either confirms of discards the disease. For the higher interval, the patient is classified as *sick* if the Ex-ECG result is positive; otherwise, the result is inconclusive and again ECHO is selected as the second test, which closes the diagnosis as *not-ill* upon a negative result or *sick* if the test is positive. For intermediate probabilities of disease, from 32% to 39% a single ECHO test suffices to reach either a *sick* or *not-ill* diagnosis, depending on the test result.

Table 4 depicts the optimal diagnosis policy, i.e. the policy which maximizes the probability of correct diagnosis. This policy is significantly more complex than the optimal cost policy in Table 3. Note that a single CTA test suffices to reach a diagnosis for a-priori probabilities in the intervals $[30\%, 34\%]$ and $[40\%, 50\%]$. For a-priori probabilities from 20% to 29% the first exam is either MRI or SPECT and reaches a diagnosis upon a negative result. A positive result of the first test is inconclusive and requires a second

15

Table 3: Optimal Cost Policy

Pre Test	First	Next Test		Total	Correct Diagnosis
Probability (%)	Test	If Positive	If Negative	Cost	Probability (%)
20	Ex-ECG	ECHO	-	R$90.39	87.90
21	Ex-ECG	ECHO	-	R$91.02	87.13
22	Ex-ECG	ECHO	-	R$91.64	87.24
23	Ex-ECG	ECHO	-	R$92.27	86.49
24	Ex-ECG	ECHO	-	R$92.90	85.62
25	Ex-ECG	ECHO	-	R$93.53	85.68
26	Ex-ECG	ECHO	-	R$94.15	85.06
27	Ex-ECG	ECHO	-	R$94.78	84.39
28	Ex-ECG	ECHO	-	R$95.41	84.16
29	Ex-ECG	ECHO	-	R$96.03	83.53
30	Ex-ECG	ECHO	-	R$96.66	83.72
31	Ex-ECG	ECHO	-	R$163.98	85.43
32	ECHO	-	-	R$165.00	84.54
33	ECHO	-	-	R$165.00	84.13
34	ECHO	-	-	R$165.00	84.39
35	ECHO	-	-	R$165.00	84.67
36	ECHO	-	-	R$165.00	83.96
37	ECHO	-	-	R$165.00	84.26
38	ECHO	-	-	R$165.00	83.95
39	ECHO	-	-	R$165.00	84.29
40	Ex-ECG	-	ECHO	R$122.07	74.30
41	Ex-ECG	-	ECHO	R$121.44	74.86
42	Ex-ECG	-	ECHO	R$120.82	75.29
43	Ex-ECG	-	ECHO	R$120.19	75.35
44	Ex-ECG	-	ECHO	R$119.56	75.35
45	Ex-ECG	-	ECHO	R$118.94	75.73
46	Ex-ECG	-	ECHO	R$118.31	75.84
47	Ex-ECG	-	ECHO	R$117.68	76.34
48	Ex-ECG	-	ECHO	R$117.05	76.85
49	Ex-ECG	-	ECHO	R$116.43	77.02
50	Ex-ECG	-	ECHO	R$115.80	77.56
51	Ex-ECG	-	ECHO	R$115.17	77.99
52	Ex-ECG	-	ECHO	R$114.55	77.73
53	Ex-ECG	-	ECHO	R$113.92	78.30
54	Ex-ECG	-	ECHO	R$113.29	78.89
55	Ex-ECG	-	ECHO	R$112.67	78.99
56	Ex-ECG	-	ECHO	R$112.04	79.60
57	Ex-ECG	-	ECHO	R$111.41	79.42
58	Ex-ECG	-	ECHO	R$110.78	80.06
59	Ex-ECG	-	ECHO	R$110.16	80.22
60	Ex-ECG	-	ECHO	R$109.53	80.89

test, CTA for a 20% a-priori probability and ECHO for the remaining values, whose result determines the diagnosis. For a-priori probabilities higher

16

than 51% the first exam is ECHO, which reaches a diagnosis if the result is positive; otherwise, a second test is selected, either MRI or SPECT, which reaches a diagnosis.

The results in Tables 3 and 4 illustrate the potential of the proposed method, which finds the best available policy with regards to the chosen performance function. Table 4 also suggests that the optimal policy, which may be readily obtained by the shortest path algorithm, may be quite complex. This contrasts with typical approaches which compare a limited set of previously designed policies, finding the best of these pre-conceived policies (e.g., Bertoldi et al., 2016; Ferreira et al., 2014). Naturally, decision makers tend to design relatively simple policies, often arbitrating a diagnosis regardless of the posterior probability after the last test. Moreover, as these methods do not evaluate all possible policies, the resulting strategy may be far from optimal. Figure 1 depicts the posterior probability of correct diagnosis for
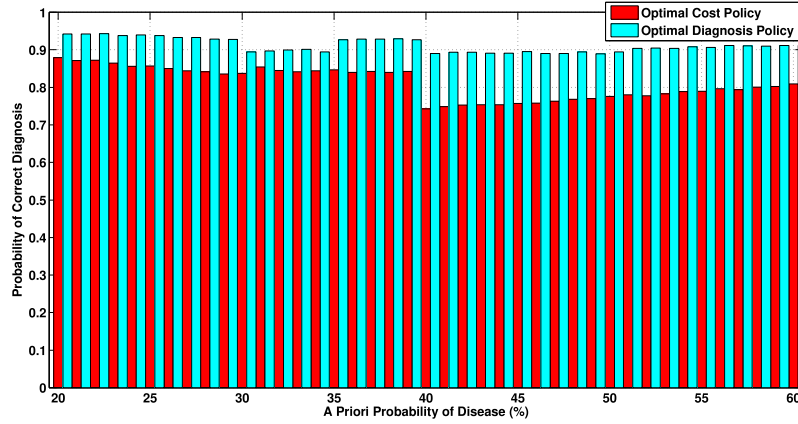

Figure 1: Correct Diagnosis Comparison

both the optimal policy with respect to overall cost and the optimal policy

17

Table 4: Optimal Diagnosis Policy

Pre Test	First	Next Test		Total	Correct Diagnosis
Probability (%)	Test	If Positive	If Negative	Cost	Probability (%)
20	MRI	CTA	-	R$482.81	94.21
21	SPECT	ECHO	-	R$928.85	94.23
22	SPECT	ECHO	-	R$930.56	94.35
23	MRI	ECHO	-	R$610.34	93.84
24	MRI	ECHO	-	R$610.24	93.95
25	MRI	ECHO	-	R$612.38	93.78
26	MRI	ECHO	-	R$612.00	93.29
27	MRI	ECHO	-	R$613.95	93.29
28	MRI	ECHO	-	R$497.51	92.84
29	MRI	ECHO	-	R$498.68	92.80
30	CTA	-	-	R$328.54	89.46
31	CTA	-	-	R$328.54	89.69
32	CTA	-	-	R$328.54	89.92
33	CTA	-	-	R$328.54	90.18
34	CTA	-	-	R$328.54	89.45
35	Ex-ECG	-	MRI	R$483.29	92.71
36	Ex-ECG	-	SPECT	R$736.39	92.89
37	Ex-ECG	-	SPECT	R$733.46	92.82
38	Ex-ECG	-	SPECT	R$731.47	92.97
39	Ex-ECG	-	MRI	R$548.13	92.67
40	CTA	-	-	R$328.54	88.99
41	CTA	-	-	R$328.54	89.36
42	CTA	-	-	R$328.54	89.33
43	CTA	-	-	R$328.54	89.15
44	CTA	-	-	R$328.54	89.13
45	CTA	-	-	R$328.54	89.56
46	CTA	-	-	R$328.54	89.00
47	CTA	-	-	R$328.54	89.00
48	CTA	-	-	R$328.54	89.46
49	CTA	-	-	R$328.54	88.94
50	CTA	-	-	R$328.54	89.44
51	ECHO	-	MRI	R$422.53	90.38
52	ECHO	-	SPECT	R$639.50	90.45
53	ECHO	-	SPECT	R$633.56	90.41
54	ECHO	-	SPECT	R$628.76	90.83
55	ECHO	-	MRI	R$413.03	90.67
56	ECHO	-	MRI	R$409.80	91.14
57	ECHO	-	MRI	R$407.63	91.10
58	ECHO	-	MRI	R$405.43	90.97
59	ECHO	-	MRI	R$403.20	91.19
60	ECHO	-	MRI	R$399.93	91.17

with respect to correct diagnosis. Unsurprisingly, the latter outperforms the former for all intermediate a priori probabilities of disease, but the differences
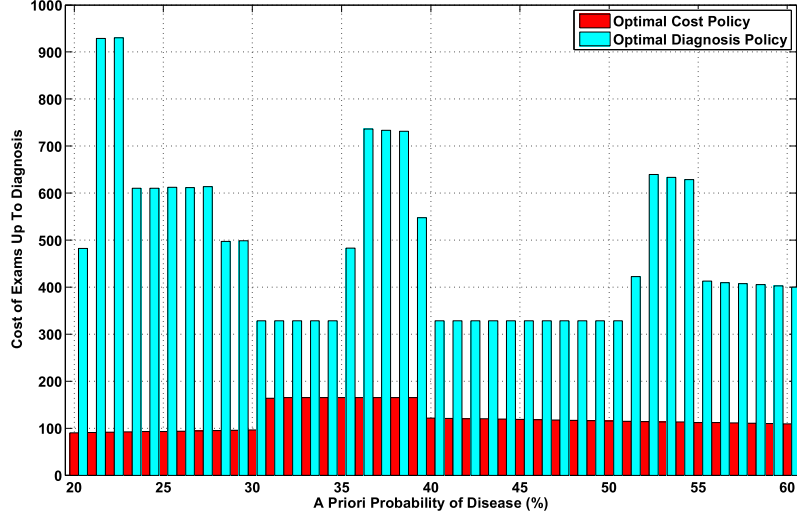
Figure 2: Cost to Diagnosis Comparison

are more pronounced in the rightmost half of the graphic. Observe that only the value x of the pair $z = (\rho, y)$ is depicted. That happens because we only depict states with $y = (0, 0, 0, 0, 0)^T$, for we are interested in the path to diagnosis of a newly arrived patient, who has taken no test so far. The same remains true for all analyses that follow.

Figure 2 features the overall costs of the optimal cost policy versus those of the optimal diagnosis policy. Note that we have a pronounced difference in costs, with larger differences for some selected values of prior probabilities.

A closer look at Figures 1 and 2 yields that the optimal diagnosis policy delivers slightly increased probabilities of correct diagnosis with respect to the most economical test strategy (policy). However, these increases demand substantial extra expenses. For a better evaluation of the costs and benefits of both strategies, Figure 3 presents the normalized cost per percentage point of
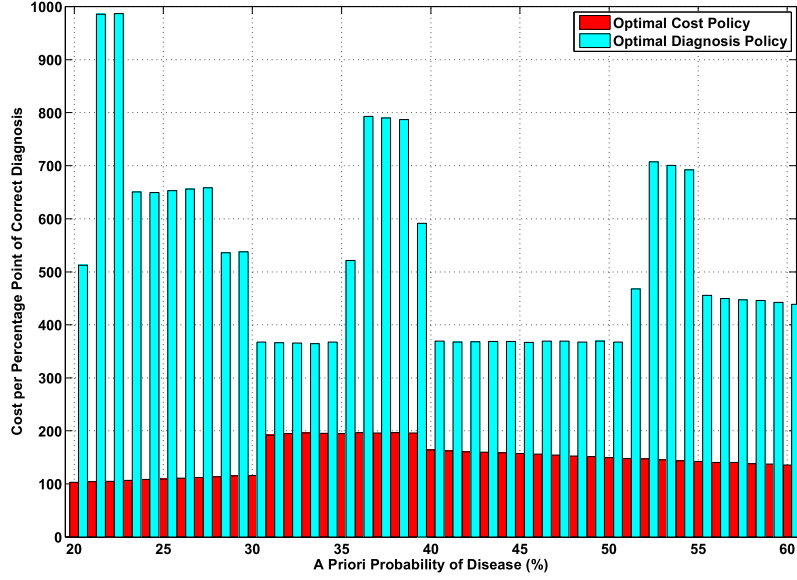
19

Figure 3: Comparison of Overall Cost per Percentage Point of Correct Diagnosis

right diagnosis for both strategies. One can see that getting a right diagnosis is much more economical by applying the optimal cost policy.

It is worth pointing out that there is no official guideline of the Brazilian government regarding the sequence of tests to be performed. However, there seems to be a consensual strategy among physicians in Brazil to take first the Ex-ECG, whenever possible, followed by either CTA or SPECT and then, if necessary, ECHO. We point out that such an strategy is similar to the strategy in Table 3 for higher probabilities of disease. However, as it prioritizes more costly tests, it is generally more expensive than the optimal cost policy in Table 3. For that reason, and also to provide a policy with the minimum number of tests, we chose to focus our analysis in the optimal diagnostic pol-

20

icy. We refer the interested reader to Appendix A for a thorough evaluation of the current practice and a comparison with the optimal strategies found in this section.

Remark 2. *It is worth pointing out that the decision maker must be aware that the physicians' estimate of the probability of the disease prior to tests is subjective, and thus may vary, possibly resulting in different test policies. In that sense, an analysis of the robustness of the policy to small variations in such an evaluation should be carried out on a case-by-case basis. With regards to the characteristics of the tests, i.e. their respective probabilities of false positive and false negative results, they do influence the outcome. However, their overall influence can be readily assessed in an analysis of the probability of correct diagnosis, such as that in Figure 1.*

3.2. Experiments with a Uncertain Likelihood Ratios and Without a Prescribed Diagnostic Region

This section explores possible generalizations of the approach. We present some ad-hoc experiments designed to illustrate some potential uses of the proposed approach. Firstly, we consider that no diagnostic region was prescribed, hence the physician would continue testing until he is satisfied with the posterior probability of disease, or until he runs out of tests to take. To help guide the system into a desirable region, we set a penalty

$$S(\rho) = -\rho \log(\rho) - (1 - \rho) \log(1 - \rho)$$

for each intermediate probability of disease ρ. Such a penalty is an entropy function (**?**), often employed in the as a measure of quality of information.

21

We set

$$c(z, a) = \alpha f(a) + (1 - \alpha)e^{S(x)}, \qquad (9)$$

where the latter term is a penalty that increases with the uncertainty of the diagnosis, while the former is the cost of taking test a. For the example, three tests are available, whose parameters appear in Table 5. It is worth pointing out that *test 3* possesses random likelihood ratios. This means that one does not know a-priori what the likelihood ratios of the test will be. In particular, the eventual likelihood ratios of an application of *test 3* are drawn from the *uniform* probability distributions $U(\cdot)$ specified in Table 5, over the intervals in parenthesis.

Table 5: Properties of the Tests.

Test	Sensitivity	Specificity	Positive Likelihood Ratio	Negative Likelihood Ratio
1	0.8	0.9	5	0.2
2	0.8	0.9	4	0.1
3	0.8	0.9	$U(4 - 5)$	$U(0.1 - 0.2)$

The costs are $f(1) = f(2) = 1.5$ and $f(3) = 0.8$. Observe that Test 1 is more precise than test 2 for positive results, whereas the opposite is true for negative results. Test 3 is less expensive than the former two, at the cost of some variation in precision. The uniform distribution implies that a positive test will result in an interval of equally likely posterior probabilities of disease, the same happening to a negative result. The idea is to emulate a

22

variation in the precision of the test results and verify the effect of it in the optimal policy.

For $\alpha = 1$, the first test of the optimal strategy, i.e. the selected test for states $z = (\rho, y) : y = [0, 0, 0]$, is selected according to the following rule:

$$\text{First test } (\alpha = 1) = \begin{cases} 1, & \text{if } 0.08 \leq \rho \leq 0.13, \\ 2, & \text{otherwise.} \end{cases} \tag{10}$$

We see that, when the focus is solely on the costs of the tests, the least expensive test is the first in the nearly all cases. The next rule defines the second test of the optimal strategy when the first test was test 3:

$$\text{Second test (after test3, } \alpha = 1) = \begin{cases} 1, & \text{if } \rho = 0.03, \\ 2, & \text{otherwise.} \end{cases} \tag{11}$$

Observe that the optimal policy assigns test 2 following test 3 whenever the probability of disease after the first test is different from $0.03(3\%)$.

Equation (12) below depicts the results for $a = 0$, which favors more precise tests:

$$\text{First test} = \begin{cases} 3, & \text{if } \begin{cases} \rho \leq 0.07 \\ 0.14 \leq \rho \leq 0.15 \\ 0.72 \leq \rho \leq 75 \\ 0.95 \leq \rho \leq 0.99 \end{cases} \\ 2, & \text{otherwise.} \end{cases} \tag{12}$$

Observe that test 2 is chosen as the first test in most cases, with test 3 being selected for some low and high a-priori probabilities. Whenever test 2

23

is selected as the first test, the optimal policy prescribes test 3 as the next. For intermediate values of α, the optimal policy generally coincides with that for $\alpha = 1$, except for very small values in the vicinity of zero, for which the optimal policy equals that of $\alpha = 0$.

4. Concluding Remarks

This paper proposes a stochastic shortest path model to find the optimal testing strategy for diagnosing a given disease. The model allows the decision maker to prescribe a testing strategy for each a-priori probability of disease, thus determining the optimal sequence of tests, which may vary depending on the results of the tests. Provided that a policy exists which warrants a diagnosis for all intermediate probabilities of disease, the optimal policy is guaranteed to reach a target set and makes sure that the posterior probability after the last test warrants an immediate diagnosis.

The proposed model is dynamic, as the decisions depend on the evolution of the a-priori probabilities after each test, and allow the decision maker to establish a benchmark policy in terms of any given performance criterion. This benchmark can be employed in the evaluation of new technologies for disease detection.

24

References

Bertoldi, E.G., Stella, S.F., Rohde, L.E., Polanczyk, C.A., 2016. Long-term cost-effectiveness of diagnostic tests for assessing stable chest pain: Modeled analysis of anatomical and functional strategies. Clinical Cardiology 39, 249–256. doi:10.1002/clc.22532.

Bertsekas, D.P., 2012. Dynamic programming and optimal control, volume 2. 4 ed., Athena Scientific, Belmont.

Cui, C., Mahajan, A., 2015. On computing optimal thresholds in decentralized sequential hypothesis testing, in: Proceedings of the IEEE Conference on Decision and Control, pp. 5284–5289. doi:10.1109/CDC.2015.7403046.

Ferreira, A.M., Gonçalves, P.A., Cardim, N., 2014. Cost-effectiveness of different diagnostic strategies in suspected stable coronary artery disease in portugal. Arquivos Brasileiros de Cardiologia 102, 391–402. doi:10.5935/abc.20140042.

Gill, C.J., Sabin, L., Schmid, C.H., 2005. Why clinicians are natural bayesians. British Medical Journal 330, 1080–1083. doi:10.1136/bmj.330.7499.1080.

Hall, G., 1967. The clinical application of Bayes' theorem. The Lancet 290, 555 – 557. doi:10.1016/S0140-6736(67)90514-4. originally published as Volume 2, Issue 7515.

Hall, G.H., Spiegelhalter, D.J., Balla, J.I., Iansek, R., Elstein, A., 1985. Diagnosis, and Bayes' theorem. The Lancet 325, 705 – 706. doi:10.1016/S0140-6736(85)91372-8.

25

Hamm, R.M., Beasley, W.H., 2014. The balance beam metaphor: A perspective on clinical diagnosis. Medical Decision Making 34, 841–853. doi:`10.1177/0272989X14528755`.

Heijenbrok-Kal, M.H., Fleischmann, K.E., Hunink, M.M., 2007. Stress echocardiography, stress single-photon-emission computed tomography and electron beam computed tomography for the assessment of coronary artery disease: A meta-analysis of diagnostic performance. American Heart Journal 154, 415 – 423. doi:`http://dx.doi.org/10.1016/j.ahj.2007.04.061`.

Lin, C.Y., Waller, L.A., Lyles, R.H., 2012. The likelihood approach for the comparison of medical diagnostic system with multiple binary tests. Journal of Applied Statistics 39, 1437–1454.

Pauker, S.G., Kassirer, J.P., 1980. The threshold approach to clinical decision making. New England Journal of Medicine 302, 1109–1117.

Pereira, B.B., Pereira, C.A.B., 2005. A likelihood approach to diagnostic tests in clinical medicine. Statistical Journal 3, 77–98.

Weintraub, W.S., Madeira, S.W., Bodenheimer, M.M., Seelaus, P.A., Katz, R.I., Feldman, M.S., Agarwal, J.B., Banka, V.S., Helfant, R.H., 1984. Critical analysis of the application of bayes' theorem to sequential testing in the noninvasive diagnosis of coronary artery disease. The American Journal of Cardiology 54, 43 – 49. doi:`10.1016/0002-9149(84)90301-1`.

Appendix A. Results for the Current Practice

This section presents the results on the probability of correct diagnosis, as well as the total expected cost up to diagnosis for the current consensual practice for diagnosing Coronary Artery Disease in Brasil. The policy, which is described at the end of Section 3.1, is detailed in Table A.6. The results

410 are further detailed in Figures A.4 and A.5.

Comparing the results in Tables 3 and 4 with those in Table A.6 , one is able to reach some conclusions regarding the strengths and weaknesses of each approach. One can observe that the current practice is more costly than the optimal cost policy and less costly than the optimal diagnosis policy, which

415 is expected. In contrast, the current practice is outperformed by the optimal diagnosis policy with respect to the probability of correct diagnosis. However, it is at least as good as the optimal cost policy in terms of probabilities of correct diagnosis.

Figure A.4 summarizes the results with respect to the probability of cor-

420 rect diagnosis. Observe that the current practice can be seen as a compromise between the two extreme policies. To evaluate the quality of the compromise, Figure A.5 conveys the cost per percentage point of correct diagnosis. One can notice that the optimal cost policy is at least as cost efficient as the current practice, while also presenting significantly better performances for

425 most of the intermediate probabilities of disease.

Table A.6: Current Consensual Practice for CAD Testing in Brazil

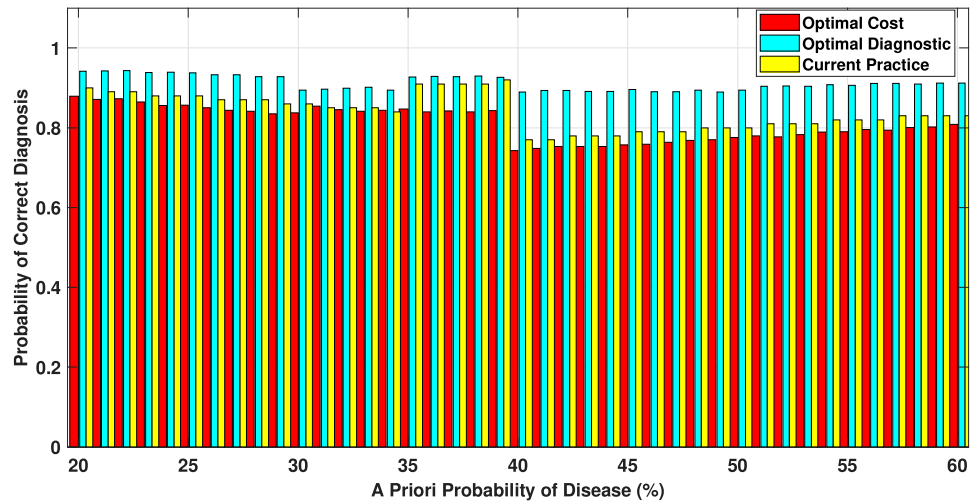| Pre Test Probability (%)	First Test	Next Test If Positive	If Negative	Total Cost	Correct Diagnosis Probability (%)
20	Ex-ECG	CTA	-	R$150,25	0,90
21	Ex-ECG	CTA	-	151,49	0,89
22	Ex-ECG	CTA	-	152,74	0,89
23	Ex-ECG	CTA	-	153,99	0,88
24	Ex-ECG	CTA	-	155,24	0,88
25	Ex-ECG	CTA	-	156,49	0,88
26	Ex-ECG	CTA	-	157,74	0,87
27	Ex-ECG	CTA	-	158,98	0,87
28	Ex-ECG	CTA	-	160,23	0,87
29	Ex-ECG	CTA	-	161,48	0,86
30	Ex-ECG	CTA	-	162,73	0,86
31	Ex-ECG	CTA	-	163,98	0,85
32	Ex-ECG	CTA	-	165,23	0,85
33	Ex-ECG	CTA	-	166,48	0,85
34	Ex-ECG	CTA	-	167,72	0,84
35	Ex-ECG	CTA	CTA	358,54	0,91
36	Ex-ECG	CTA	CTA	358,54	0,91
37	Ex-ECG	CTA	CTA	358,54	0,91
38	Ex-ECG	CTA	CTA	358,54	0,91
39	Ex-ECG	CTA	CTA	358,54	0,92
40	Ex-ECG	-	CTA	213,33	0,77
41	Ex-ECG	-	CTA	212,08	0,77
42	Ex-ECG	-	CTA	210,83	0,78
43	Ex-ECG	-	CTA	209,58	0,78
44	Ex-ECG	-	CTA	208,33	0,78
45	Ex-ECG	-	CTA	207,08	0,79
46	Ex-ECG	-	CTA	205,83	0,79
47	Ex-ECG	-	CTA	204,59	0,79
48	Ex-ECG	-	CTA	203,34	0,80
49	Ex-ECG	-	CTA	202,09	0,80
50	Ex-ECG	-	CTA	200,84	0,80
51	Ex-ECG	-	CTA	199,59	0,81
52	Ex-ECG	-	CTA	198,34	0,81
53	Ex-ECG	-	CTA	197,10	0,81
54	Ex-ECG	-	CTA	195,85	0,82
55	Ex-ECG	-	CTA	194,60	0,82
56	Ex-ECG	-	CTA	193,35	0,82
57	Ex-ECG	-	CTA	192,10	0,83
58	Ex-ECG	-	CTA	190,85	0,83
59	Ex-ECG	-	CTA	189,60	0,83
60	Ex-ECG	-	CTA	188,36	0,83

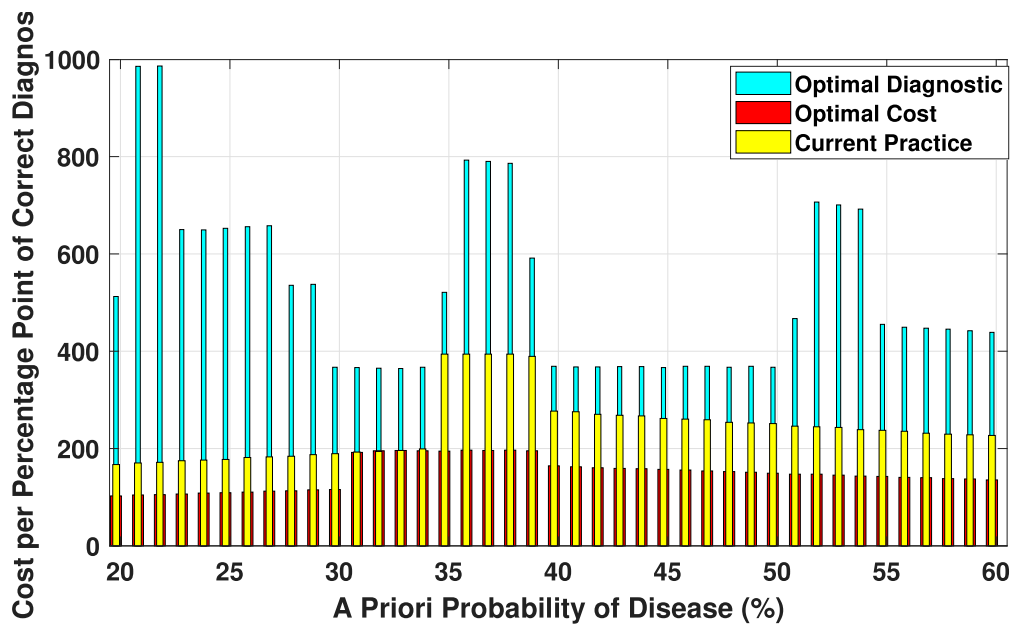Figure A.4: Correct Diagnosis Comparison, Including Current Practice

Figure A.5: Comparison of Overall Cost per Percentage Point of Correct Diagnosis, Including Current Practice

29