Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

Journal of Evaluation in Clinical Practice ISSN 1365-2753

Towards theory integration: Threshold model as a link between signal detection theory, fast-and-frugal trees and evidence accumulation theory

Iztok Hozo PhD,[1] Benjamin Djulbegovic MD PhD,[3,4,5] Shenghua Luan PhD,[6] Athanasios Tsalatsanis PhD[2] and Gerd Gigerenzer PhD[5]

[1]Professor, Department of Mathematics, Indiana University, Gary, IN, USA
[2]Associate Professor, USF Health Program for Comparative Effectiveness Research, Division for Evidence-Based Medicine, Department of Internal Medicine, University of South Florida, Tampa, FL, USA
[3]Professor of Oncology, Departments of Hematology and Health Outcome Behavior, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
[4]Director of Medical Research, Tampa General Hospital, Tampa, FL, USA
[5]Professor, [6]Researcher, Max Planck Institute for Human Development, Berlin, Germany

Keywords
clinical guidelines, evaluation, evidence-based medicine

Correspondence
Professor Benjamin Djulbegovic
12901 Bruce B. Downs Blvd, MDC02
Tampa, FL 33612
USA
E-mail: bdjulbeg@health.usf.edu

Accepted for publication: 6 November 2015

doi:10.1111/jep.12490

Abstract

Rationale, aims and objectives Theories of decision making are divided between those aiming to help decision makers in the real, 'large' world and those who study decisions in idealized 'small' world settings. For the most part, these large- and small-world decision theories remain disconnected.

Methods We linked the small-world decision theoretic concepts of signal detection theory (SDT) and evidence accumulation theory (EAT) to the threshold model and the large world of heuristic decision making that rely on fast-and-frugal decision trees (FFT).

Results We connected these large- and small-world theories by demonstrating that seemingly different decision-making concepts are actually equivalent. In doing so, we were able (1) to link the threshold model to EAT and FFT, thereby creating decision criteria that take into account both the classification accuracy of FFT and the consequences built in the threshold model; (2) to demonstrate how threshold criteria can be used as a strategy for optimal selection of cues when constructing FFT; and (3) to show that the compensatory strategy expressed in the threshold model can be linked to a non-compensatory FFT approach to decision making. We also showed how construction and performance of FFT depend on having reliable information – the results were highly sensitive to the estimates of benefits and harms of health interventions. We illustrate the practical usefulness of our analysis by describing an FFT we developed for prescribing statins for primary prevention of cardiovascular disease.

Conclusions By linking SDT and EAT to the compensatory threshold model and to non-compensatory heuristic decision making (FFT), we showed how these two decision strategies are ultimately linked within a broader theoretical framework and thereby respond to calls for integrating decision theory paradigms.

Introduction

Theories of decision making deal with either 'large-' or 'small'-world phenomena [1]. In a small world, decision makers are not under time pressure and have access to all relevant knowledge, including all alternatives, consequences and probabilities. In turn, such knowledge enables a decision maker to make an optimal, rational decision. A prototype of 'small'-world theory is expected utility theory (EUT). By contrast, in a 'large' (i.e. typically a real) world, knowledge about the complete set of alternatives, consequences and probabilities is limited. Under these circumstances, a rational decision maker relies on adaptive cognitive processes that surprisingly often lead to efficient and accurate choices. A prototype of large-world theory is the heuristic theory of decision

making [2]. Until lately, no attempt had been made to connect decision theories of large worlds with those of small ones. Recently, however, Luan *et al*. successfully applied small-world signal detection theory (SDT) to fast-and-frugal tree (FFT) heuristics [3]. In a separate undertaking, Lee and Cummins [4] proposed evidence accumulation theory (EAT) as a concept unifying heuristics and 'rational' models.

In a further endeavor to unify different psychological and statistical theories [5], we link the threshold model (formulated within both the EUT and regret framework) [6–8] with SDT and FFT and with EAT. We illustrate the usefulness of this new framework in the context of medical decision making with the example of prescribing statins for primary prevention of cardiovascular disease (CVD).

Methods

SDT

SDT, which has its origins in the Neyman–Pearson theory of hypothesis testing, is widely used throughout science and medicine [3]. Its fundamental assumption resides on the notion that the two possible events (*signal*, e.g. presence of disease, and *noise*, e.g. absence of disease) have overlapping distributions on an observation scale X [9]. Each of these distributions is further divided into two possible outcomes, which are determined by setting a decision criterion (x_c, as in Fig. 1). The criterion divides the signal distribution into true positives (hits) and false negatives (misses). In medicine, the hit rate is called *sensitivity* (S) and

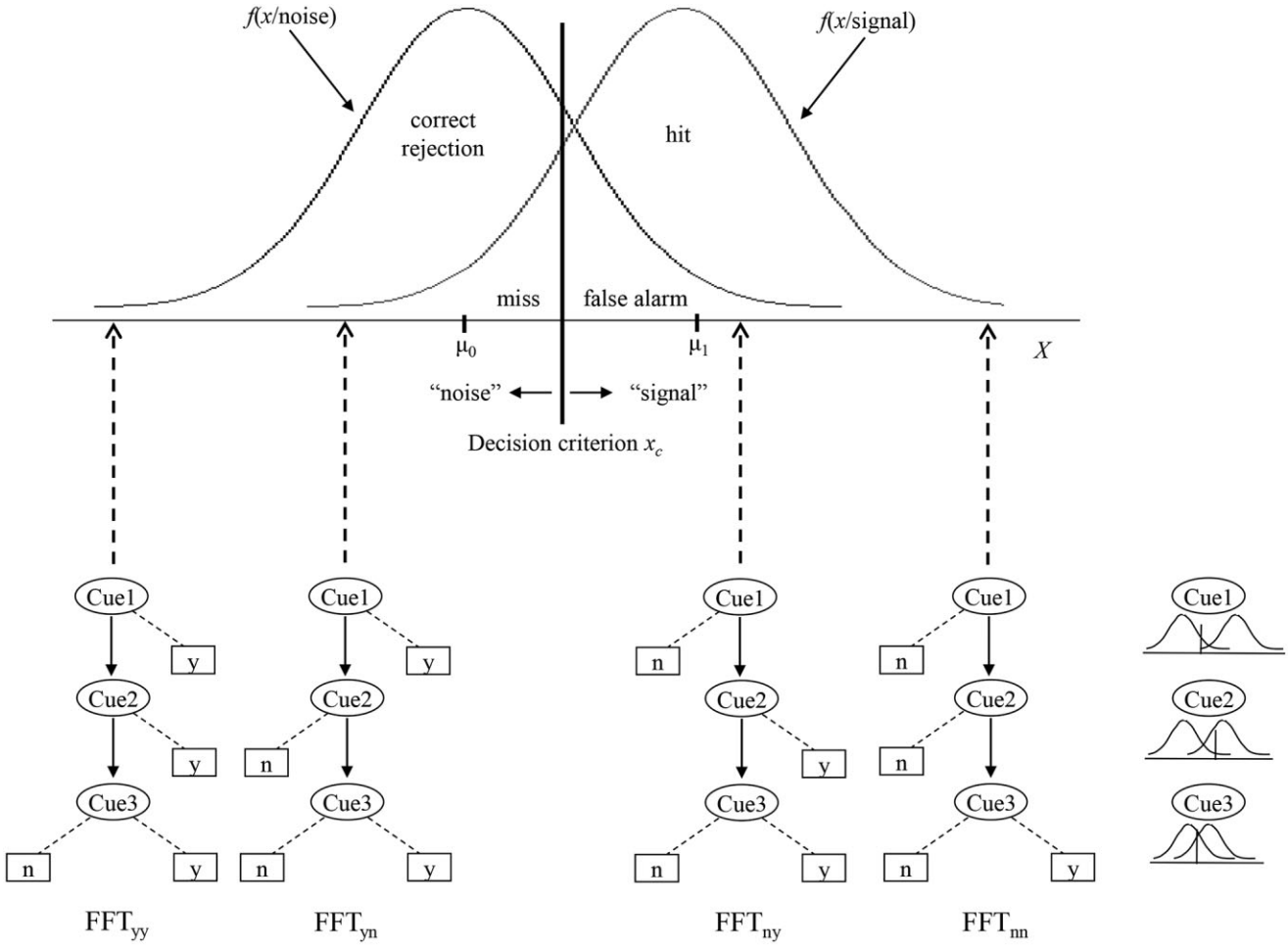

Figure 1 A relationship between signal detection theory (SDT) and fast-and-frugal trees (FFT). The upper part of the figure illustrates the concepts of SDT in a binary decision task, and the lower part illustrates the four possible FFTs that can be constructed when three cues are searched in a set order. Based on the decisions pointed to by the first two exits, the trees are named from left to right FFTyy, FFTyn, FFTny and FFTnn (where y stands for 'yes' and n for 'no'). The arrows connecting the figure parts indicate the rough locations of the four FFTs' decision criteria when they are used to make a binary y/n (for signal and noise, respectively) decision. Among the four, FFTyy has the most liberal decision criterion, and FFTnn the most conservative one. The decision criteria of FFTyn and FFTny are less extreme than the other two, with FFTyn being more liberal than FFTny. The two overlapping normal distributions next to each cue illustrate SDT's assumption of how object values are distributed on a cue and emphasize that each cue comes with its own discriminability and decision criterion (in *d'* and *c*, respectively; see Table 1 – Appendix). Note that the structure of FFTs shown in the figure is not based on contrasting it to the threshold (the latter is shown in Fig. 5; see text for further explanations; figure is based on Luan *et al.* [3]).

1 – sensitivity is the *false-negative rate* (FN). The noise distribution is composed of true negatives (correct rejections) and false positives; their rates are called *specificity* (C) and *false alarm rate* (FA), respectively [9]. From these data, we can calculate the accuracy of classification of a particular diagnostic cue. Because the consequences of misses versus false alarms often differ, two additional metrics are typically calculated according to SDT. These are *d'* (discrimination) and *c* (decision criterion) [9]. Discrimination or discriminability (*d'*) measures the distance between the means of signal and noise in standard deviation units (*z*-scores). The decision criterion (*c*) (also known as *response bias* [3,9]) determines the decision cut-off; if it is set at $c = 0$, then *FA* and *FN* are weighted according to the prior probabilities of signal and noise. If *c* is moved to the left (<0), then $FN < FA$ relative to prior probabilities of signal and noise (*liberal bias*). The opposite holds when *c* is moved to the right (>0). In this case, *FN*s are more tolerated than *FA*s ($FN > FA$), relative to prior probabilities of signal and noise (*conservative bias*). The threshold likelihood ratio of the decision at hand according to SDT is given as:

$$\beta_{optimal} = \frac{V(C) - V(FA)}{V(S) - V(FN)} \cdot \frac{p(Noise)}{p(Signal)} \quad (1)$$

where $V(\cdot)$ represents the utility of a particular outcome [3] (see section on threshold model below). For terminology of SDT, see Appendix A.

FFTs

FFTs are a class of simple heuristic decision-making strategies that relies on limited information to reduce estimation error and facilitate fast decisions [1,2,10]. An FFT is a decision tree composed of sequentially ordered cues. Typically, cues and decisions are binary (yes/no), and their relation can be framed as *if–then* statements (e.g. if a person has severe chest pain, then perform diagnostic tests to rule out myocardial infarction). If the condition is met, the decision can be made and the FFT is exited. If the condition is not met, the FFT considers the other cues, one after another, until the exit condition of a cue is met. The last cue of an FFT has two exits to ensure that a decision is ultimately made [3]. Formally, an FFT is defined as a decision tree that has $m + 1$ exits, with one exit for each of the first $m - 1$ cues and two exits for the last cue [3].

An FFT relies on the so-called *non-compensatory* decision making; once the tree is exited, cues lower in the decision tree hierarchy cannot compensate for cue information higher in the hierarchy [1,11]. Surprisingly, by ignoring information, FFTs can be more accurate than statistical multivariate regression models ('less is more') [2]. This is because FFTs can be less susceptible to overfitting than the regression models [2].

Every cue in a FFT can correctly or incorrectly classify signal and noise. The performance of each cue can be described using SDT criteria [3]. Figure 1 illustrates the application of SDT to FFT. Thus, decision theories developed for small worlds can be directly linked to large-world decision theories [1,2]. The most important aspect of Fig. 1 is that the exit structure of the FFTs determines the ratio between false negatives and false positives. For example, the structure FFTyy on the far left side of Fig. 1 has a high hit rate (sensitivity) at the expense of a large rate of false alarms. The FFTnn on the far right side of the figure reflects the opposite: the FA is reduced at the expense of a large rate of false negatives.

EAT

Whereas SDT is based on Neyman–Pearson statistics, EAT is based on Abraham Wald's sequential decision theory, which assumes that decisions are made by accumulating evidence via a sequential sampling process [3]. Lee and Cummins showed [4] that sequential heuristics such as take the best (TTB) can be related to traditional 'rational' approaches to decision making by introducing an action threshold. TTB is a lexicographic heuristic for deciding between two objects, similar in spirit to an FFT, which assigns one object to one of two categories. Decisions in the framework of heuristic decision making often occur after the first piece of evidence is deemed sufficient for action (according to the stopping rule of TTB, similar to that of an FFT), whereas traditional rational approaches such as EUT require all of the available information to be sampled before a decision is made. Lee and Cummins [4] argued that 'by setting different threshold levels of evidence required for decision making, both the heuristics decision making and the rational models become special cases of a more general evidence accumulation account'.

Threshold model

A popular benefit–risk analysis in clinical medicine has relied on the application of the threshold model, which has been developed both in EUT and expected regret framework [6–8]. As noted earlier, in this paper we describe a signal in terms of a patient having a disease with some probability, *p*. According to the threshold model, when a physician is faced with uncertainty about whether to treat or simply further observe the patient, there must exist some *p* at which prescribing or not prescribing treatment are equal options (see Fig. 2) [6–8]. This probability is referred to as the threshold probability (*T*) [6–8]. Analysing the tree shown in Fig. 2, we find that the physician should prescribe treatment if *p* is

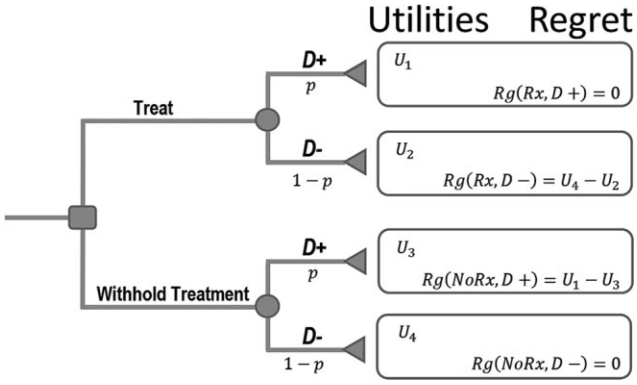

Figure 2 A decision tree from which the threshold model is derived. A physician has two choices: treat (e.g. prescribe statins) versus do not treat. The patient may or may not have disease (D) with the probability *p*. The probability at which a decision maker is indifferent between acting (e.g. treating) or not acting (e.g. withholding treatment) is called the threshold probability (see Appendix B for details).

larger than T and should withhold treatment if p is less than T. Figure 2 shows a decision tree where outcomes are expressed both in terms of utility and in terms of regret. Detailed derivations of the inequalities above are shown in Appendix B. Note that the regret model and EUT produce the same results as long as regret is a linear function of lost potential utilities, as is often the case in clinical settings [12,13].

One important aspect of the tree in Fig. 2 concerns the definition of treatment benefits and harms. In the original threshold model, these were defined as net benefits (B) and net harms (H). B represents the difference in the utility of the outcomes if a patient *with* disease was treated versus not treated; H is defined as the difference in the utility of the outcomes for a patient *without* the disease. That is, B = utility of true positives – utility of false negatives ($B = U_1 - U_3$ in Fig. 2); H = utility of true negatives – utility of false positives ($H = U_4 - U_2$ in Fig. 2) [6–8], similar to the classification of decision errors according to SDT.

This brief description of the threshold model lends itself to outlining a rationale for application of the threshold model to FFTs: a cue in an FFT can be selected according to the threshold model. Appendix B provides a brief derivation of the threshold model and Appendix C a side-by-side comparison of the threshold model with SDT.

Linking threshold model with SDT and FFT

As stated earlier, according to the threshold model, treatment is withheld if the posterior probability of disease is smaller than $\frac{1}{1+\frac{B}{H}}$ and is prescribed if the probability is larger than $\frac{1}{1+\frac{B}{H}}$.

Appendix C demonstrates that $\beta_{optimal}$ (Eq. 1) is equivalent to the action threshold in the threshold model.

Despite this mathematical equivalence between SDT and the threshold model, as demonstrated further below, the explicit linkage of the threshold to SDT (via EAT) can change the structure of an FFT. This, in turn, may affect both classification accuracy and decision consequences. More importantly, under some circumstances, using the threshold to calculate SDT statistics may not necessarily result in an orderly increase in c statistics from the left (FFTyy) to the right (FFTnn), as in the case of standard FFTs shown in Fig. 1. To distinguish it from a standard FFT, we henceforth refer to the threshold-derived FFT as an FFTT.

As noted, as long as the relationship remains linear, identical results can be derived using regret theory framework [12,13]. However, if preferences are elicited using regret scales [14], the differences between EUT and regret threshold model are often dramatic, even though their mathematical formulation is identical [8,15].

In the next three sections, we (1) show an application of the threshold model to both aid selection of cues for generating FFTs and help make better decisions; (2) describe the linking of the threshold model to EAT; and (3) provide a practical illustration of the approach outlined in this paper by focusing on decision differences between FFT and FFTT.

Selection of cues for FFT

Martignon and colleagues described two strategies to select cues for FFTs: MAX and ZIG (or zigzag, dual max) [16]. These strategies entail selecting a cue with the highest predictive accuracy but do not consider the consequences of a right or wrong decision.

The threshold model can be applied to determine when each cue should be considered *before* it is actually searched (Fig. 3).

The following relation holds for the situation before each cue is considered

1 Whenever the positive likelihood ratio ($LR+$) is *smaller* than $\beta_{optimal}$, we should not consider the cue (we should accept noise, or consider a new cue)

2 Whenever the negative likelihood ratio ($LR-$) is *larger* than $\beta_{optimal}$, we should not consider the cue (we should accept signal, or consider a new cue)

3 If both $(LR+) > \beta_{optimal}$ and $(LR-) < \beta_{optimal}$, we should proceed with considering a given cue and refine likelihood ratios accordingly.

Hence, the threshold model effectively determines the point when searching for more information will not change our decision. This is a critical aspect of effective application of heuristics: deciding when to stop further search, which, if continued, may prove to be counterproductive, costly or even detrimental [2].

It can be particularly challenging to determine a cue's cut-off when the cue is continuous. In such cases, the threshold model can be helpful. It can be shown that the optimal cut-off occurs at the point on a Receiver Operating Characteristics (ROC) curve plotting true positives (S) versus false positives (1 – C) where its slope is given by [17]:

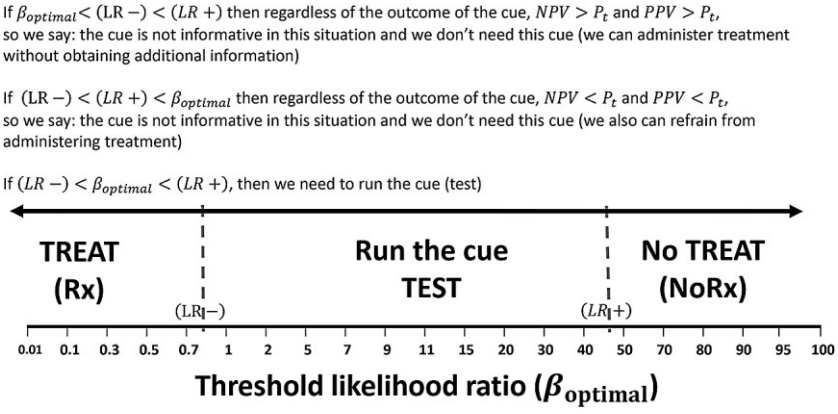

Figure 3 Selecting a cue according to the threshold model. Note how the threshold model effectively determines the point when search for more information will not change the decision. This is a critical aspect of effective application of heuristics: deciding when to stop further search, which, if continued, may turn to be counterproductive, costly or even detrimental (see text for details).

$$slope\ of\ ROC\ curve = \frac{H}{B} * \frac{p(Noise)}{p(Signal)} \quad (2)$$

Threshold model and EAT

As described earlier, EAT has been proposed as a unifying theory that can account for both heuristic decision making (such as TTB) and rational, EUT models [4]. Here we show how the threshold model can further unify SDT, FFT and EAT within EUT and the expected regret format. Lee and Cummins proposed that decision making follows sequential accrual of information to allow for comparison between two stimuli (choices) [4] and that when evidence exceeds a threshold, a decision is made. However, this threshold is never formally defined. Here we show that the threshold is equal to $\ln\beta_{optimal}$. The decision should be made when the (ln) of the sum of existing evidence $\geq\ln\beta_{optimal}$ according to

$$\sum_{a_i=1} \ln(LR_i+) + \sum_{a_i=0} \ln(LR_i-) \leq \ln\beta_{optimal} \quad (3)$$

In other words, for each patient with a sequence of cue answers (a_1, a_2, . . . , a_k) we can evaluate the sum of the logs of the likelihood ratios and see whether it is smaller than the log of $\beta_{optimal}$. If the left-hand sum is smaller than $\ln\beta_{optimal}$, treatment should be withheld, and if it is larger, treatment should be prescribed. Appendix D shows a formal linkage between EAT and the threshold model. As illustrated below, Eq. (3) allows treatment to be individualized for people with different cue values; it directly takes the consequences of treatment into account, whereas a standard FFT focuses on the accuracy of classification. That is, determining the accuracy statistics (*d'* and *c*) based on the combination of cues' true positives and true negatives in standard FFT–SDT format may not necessarily be equivalent to the classification based on the consequences of treatment according to Eq. (3) (see below for the specific illustration of the difference in the results between standard FFT and FFTT).

An illustration

We now illustrate linking these four theories – threshold, SDT, FFT and EAT – in the setting of prescribing statins for primary prevention of CVD. Statins are promoted as effective drugs for both primary and secondary prevention of CVD, the leading cause of mortality and morbidity in the United States and in most other economically developed nations [18,19]. Currently, the American College of Cardiology and the American Heart Association (ACC/AHA) recommend statins if the 10-year risk of myocardial infarction or stroke is ≥7.5% [20]. The risk of CVD can be estimated using multivariate regression models such as Framingham Risk Score (FRS) [21], derived from the Framingham Heart Study cohort [22]. However, many doctors find the use of the FRS and similar predictive models cumbersome and do not rely on them. In addition, the new guidelines recommend that doctors should not make their decisions based on laboratory values such as cholesterol levels [23]. As a consequence, most physicians rely on their clinical judgments and intuitive heuristics to advise their patients about taking statins. The important question is how formally defined heuristics such as FFT fare against multivariate regression models such as FRS. To answer this question, we obtained the data from the Framingham cohort from the National Heart, Lung, and Blood Institute in order to compare the accuracy of inferences based on FFT aided by SDT and threshold models with those based on the FRS model. Note, however, that our main goal was not to develop a new tool (FFT) that will replace decades of epidemiological and clinical research but instead to illustrate how new clinical tools can be developed using simple adaptive tools such as FFT while retaining a coherence with statistical decision theories.

The original data include the following eight variables: age, gender, total cholesterol, high-density cholesterol, systolic blood pressure, whether the patient received treatment for hypertension, whether the patient smoked and whether the patient had diabetes. All these variables were statistically associated with CVD over 10 years. Our interest was in generating an easy-to-use FFT containing ideally no more than three to five variables in order to retain its clinical usefulness. Given that new ACC/AHA guidelines de-emphasize laboratory measurements [20], we were particularly interested in generating an FFT based on clinical variables only. We aimed to compare the FFT with the FRS model at the recommended threshold of ≥7.5% of CVD [20].

To select a cue, we determined $\beta_{optimal}$ and compared it with the threshold LRs for each cue (Fig. 3). If the rule was not met, a cue was not selected. Table 1 shows the performance criteria of eight cues that comprise the FRS model in comparison with our selection criteria threshold rule shown in Fig. 3. The key determinants for the cue selection are benefits (B) and harms (H) of statins and the probability of CVD. The benefits and harms of statins in the setting of primary prevention of CVD are debatable. A Cochrane systematic review of four randomized trials enrolling 35 254 patients found a statistically significant effect of statins on two health outcomes only: statins reduced CVD by 1.33–2.5 percentage points in terms of absolute risk at the expense of an absolute increase of 0.4 percentage points in incidence of diabetes [19]. Other authors, however, estimated benefits of statins to be around 1% and incidence of diabetes about 1% [24]. In addition, statins have also been reportedly associated with a number of other harms (such as myalgia, liver test abnormalities, rhabdomyolysis) [25] that have not been precisely quantified in control trials [19] but that reduce the B/H ratio. Because of the uncertainty regarding the actual B/H of statins in this setting, there is an intense debate in the literature whether statins are more beneficial or more harmful to patients for primary prevention of CVD [23,26–28]. To take these uncertainties into account, we selected cues for the range of plausible B/H for primary prevention of CVD from as low as 0.5 to as high as 6. To make the FRS data more relevant, we assumed the baseline prevalence for CVD of 6% as the current best estimate of CVD in the general population [29].

From Table 1, it can be seen that across the wide range of B/H, three cues met our criteria for inclusion in an FFT: age, whether the patient has been diagnosed with diabetes and whether the patient is being treated for high blood pressure. Notably, these cues are always considered during a typical patient–physician encounter. Assuming that the B/H ratio of statins for primary prevention is between 1.33 and 2.5 [19], we determined the cut-off for age to be 39–49 years, with the best estimate at 44. Because our goal is to illustrate methods and not necessarily to generate a new tool for clinical practice, we report only analysis based on a B/H ratio of 2.5.

Table 1 Selection of cues to create a fast and frugal decision tree (FFT) for administration of statins in prevention of cardiovascular disease (CVD)

	Statins' benefit/harms (B/H) ratio*; pCVD = 0.25 and pCVD = 0.06 (bold)							
Cues	≤0.1[§]	0.2–0.6	0.7–1.1	1.2–1.6	1.7–2.1	**2.2–4**	4 to <5	≥5[¶]
Age (≥44)				**YES**	**YES**	**YES**	YES	
Gender						**YES**	YES	
Smoker								
Diabetic		YES	YES	**YES**	**YES**	**YES**		
Blood pressure treated			YES	**YES**	**YES**	**YES**		
Systolic blood pressure (≥125 mmHg)					YES	YES	YES	
Total cholesterol (≥250 mg dL^{-1})[†]					YES[‡]	YES	YES[‡]	
HDL cholesterol (≥40 mg dL^{-1})					YES	YES		

Bold texts refer to cues selected for the FFT described in this manuscript.

*At probability of cardiovascular disease (pCVD) of 6% (general population prevalence).

[†]If cut-off is selected at 220 mg dL^{-1} then total cholesterol can also be selected for B/H 1.2–1.6.

[‡]At cut-off of 235 mg dL^{-1}.

[§]No cue should be selected (completely uninformative) ($\beta_{optimal} > LR+$) (see Fig. 3).

[¶]No cues should be selected, that is, all patients should be treated ($\beta_{optimal} < LR-$) (see Fig. 3).

HDL, high-density lipoprotein.

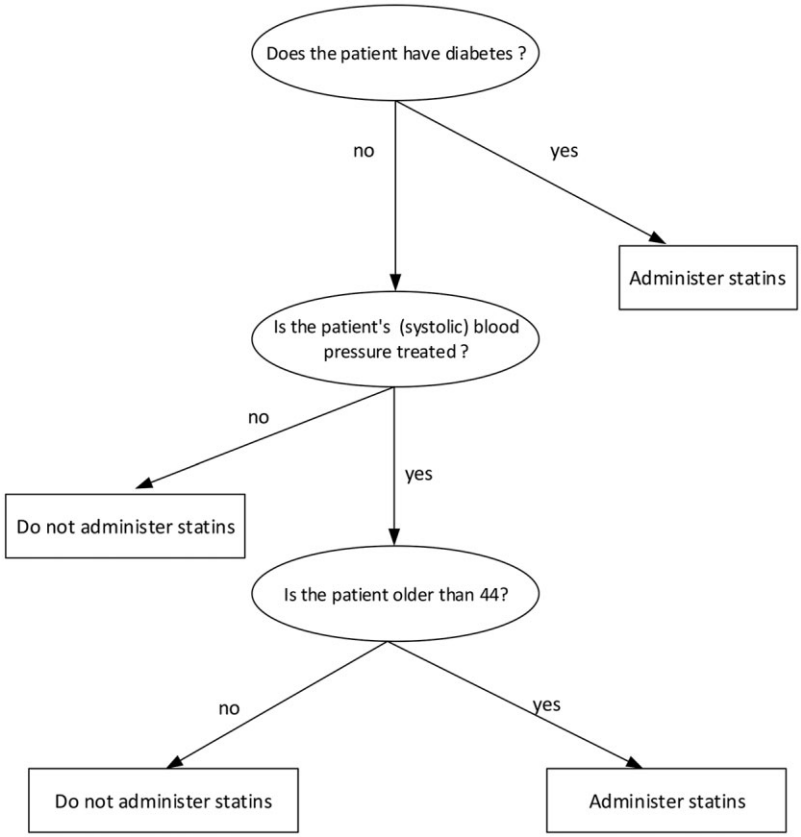

Figure 4 Fast-and-frugal tree (FFT) for prescribing statins for primary prevention of heart disease. An example of an FFTyn structure (see Fig. 1). The FFT was generated based on selection of three cues obtained from the initial patient visit: Does the patient have diabetes (yes/no)? Is the patient being treated for high pressure (yes/no)? Is the patient older than 44 (yes/no)? (see also Table 1). Note that because the last (in this case, third) cue has both exits, the FFT is identified by the cue exits before the final one (e.g. FTTyy, FFTyn, FFTny, FFTnn).

Figure 4 shows an FFT that we generated based on selection of these three cues obtained from the initial patient visit: Does the patient have diabetes (yes/no) (Y/N)? Is the patient being treated for high blood pressure (Y/N)? Is the patient older than 44 (Y/N)? The FFT stipulates that a physician will prescribe statins whenever the patient exits the tree structure after answering 'yes' and withhold prescribing statins whenever the patient exits the tree structure after answering 'no'. Therefore, there are four possible FFTs and for each individual patient there are eight possible exit paths: NNN, NNY, NYN, NYY, YNN, YNY, YYN, YYY (see also Figs 1 & 5).

The assumed B/H ratio (Table 1), however, gives equal weight to adverse effects of statins such as developing diabetes, muscle pain and liver toxicity. Because people may weigh these clinical

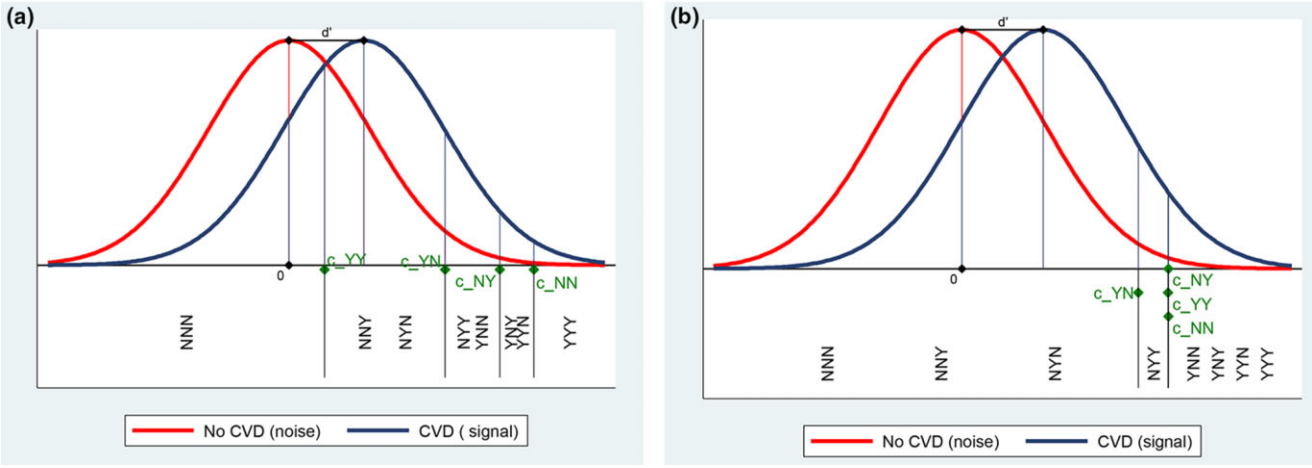

Figure 5 A relationship between fast-and-frugal trees (FFTs) with two cues and all possible paths that a patient can go through (denoted by a three-letter combination below the *x*-axis). Discriminability (*d'*) and decision criterion statistics (here denoted as *c*) are shown for each FFT first without considering the threshold (standard FFT – panel a) and where decision making depends on the threshold (FFTT – panel b). Treatment is indicated for all patients in the FFT paths to the right of the decision criterion (vertical lines passing through the green dots denote decision criterion *c*). Note, however, that in the case of FFTs (without a threshold) the signal (CVD = cardiovascular disease) significantly overlaps with noise (no CVD) for FFTyy, which means that the classification capacity of this FFT is likely unreliable. However, the structure of FFTyn, FFTny and FFnn shifts the decision criterion to the right with a much smaller overlap between the signal and noise. Here, all patients who exit on a 'yes' branch will receive treatment (denoted as the last letter 'y' in the FFT path), and all patients who exit on a 'no' branch will not receive treatment (denoted as the last letter 'n' in the FFT path). On the other hand, for FFTTs, the threshold (or in this case $\ln\beta_{optimal}$) can override the FFT exit decision, and treatment is sometimes withheld (or prescribed) for a patient who may have exited on a 'yes' (or 'no') exit. For example, as shown in panel (b), in an FFTT it is possible to treat patients along the path defined by FFTyn who would not be treated in an FFTyy. This is not possible in a standard FFT (a). (See Table 3 for comparison of treatment decisions between the FFT, FFTT and FRS models.)

outcomes differently, we also conducted a sensitivity analysis by assigning diverging weights to the failure to prevent CVD (regret of omission) and to incurring adverse events (regret of commission).

To test for the differences between the performance of FFTT (which applies the threshold model to derive classification and decision criteria) versus FFT (standard FFT, which refers to no threshold for making classification and decision recommendations) versus the FRS model (according to which statins should be prescribed if the probability of CVD is ≥7.5%), we also conducted a formal hypothesis test using the approach described by Wickens [30]. The null hypothesis of no difference was rejected if the probability of observing *z*-statistics was ≤0.05 [30].

Results

Table 2 shows the performance metrics of FFTT, FFT and the FRS model. Note that FRS performance characteristics remain identical, and that performance of FFFT does not always change predictably across all FFTT combination of cues. Unlike for FTT, where *c* statistics increase from FFTyy to FFTnn, *c* statistics are unpredictable for FTTT. That is because the threshold value is such that the balance of benefits and harms of statins remains constant across the combinations of cues and – depending on a given path – it is possible that more patients can be treated, for example, with FFTyn than with FFTyy (see Table 3 and Fig. 5). In general, FFFT retained high specificity across the cues, but at expense of very low sensitivity. The same held for all FFTs except for FFTyy, where specificity was more modest (80%) and sensitivity was higher, albeit not at a level to be considered informative (49%).

With respect to *d'* (discrimination, arguably the most important measure in SDT), FFTT and FRS performed better than FFT. Discriminability of FFTTyy was superior to FFTyy ($P = 0.0078$), whereas no difference was detected between FFTTyy and the FRS model ($P = 0.61$); the FRS model was, however, superior to FFT ($P = 0.003$). At the same time, no difference was detected between FFTTnn versus FFTnn and the FRS model ($P = 0.412$ for FFTTnn versus FFTnn; $P = 0.60$ for FFTnn versus FRS; $P = 0.76$ for FFTTnn versus FRS).

Overall, the performances of both FFT and the FRS model were far from perfect. This is not particularly surprising in light of the recent external validation of four most popular CVD risk prediction models (including FRS) demonstrating that all existing models fall short of being highly predictive of CVD [31,32]. All the models tested had high and positive values for *c*. The *c* value for FFTTyy was higher than for FFTyy ($P = 0.00001$) and FRS ($P = 0.00001$) but was identical for FFTTyn and FFTyn (both had a larger *c* value than the FRS model, at $P = 0.00001$). For FFTny and FFTnn, the *c* value was larger than that of their respective FFTTs (at $P = 0.00001$). Both *c* values remain larger than that of FRS ($P = 0.00001$).

In general, a positive *c* value means that more weight is placed on avoiding false positives, while a negative *c* value indicates a preference for avoiding not treating someone with CVD (i.e. avoiding false negatives). In light of the current controversy about B/H of statins and the estimation that more than a billion people worldwide would be prescribed the treatment if the AHA/CCA guidelines were adhered to [28], it is important to identify a simple tool for avoiding unnecessary treatment. That is, more important than the accuracy statistics shown in Table 2 are the decision

Table 2 Performance of FFTT versus FFT versus FRS* model[†§]

	FFT (YY) with threshold	**FFT (YY) w/o threshold**	**FRS (EUT = 0.075)[‡] model**
Sensitivity	0.0197 (0.0196, 0.0197)	0.4934 (0.4934, 0.4935)	0.4348 (0.4347, 0.4348)
Specificity	0.9989 (0.9989, 0.9989)	0.8074 (0.8073, 0.8075)	0.8856 (0.8856, 0.8857)
LR+	18.3996 (5.4321, 62.3226)	2.5621 (2.3178, 2.8321)	3.8013 (3.3502, 4.3132)
LR–	0.9814 (0.9724, 0.9905)	0.6274 (0.5871, 0.6705)	0.6382 (0.6021, 0.6765)
c	2.5658 (2.5624, 2.5692)	0.4424 (0.4416, 0.4432)	0.6839 (0.6832, 0.6846)
d'	1.0096 (1.0033, 1.0159)	0.8520 (0.8471, 0.8568)	1.0394 (1.0349, 1.0438)
	FFT (YN) with threshold	**FFT (YN) w/o threshold**	**FRS (EUT = 0.075) model**
Sensitivity	0.0677 (0.0676, 0.0678)	0.0677 (0.0676, 0.0678)	0.4348 (0.4346, 0.4349)
Specificity	0.9915 (0.9915, 0.9915)	0.9915 (0.9915, 0.9915)	0.8856 (0.8856, 0.8857)
LR+	7.9220 (4.9748, 12.6153)	7.9220 (4.9748, 12.6153)	3.8013 (3.3502, 4.3132)
LR–	0.9403 (0.9338, 0.9572)	0.9403 (0.9238, 0.9572)	0.6382 (0.6021, 0.6765)
c	1.9390 (1.9374, 1.9406)	1.9390 (1.9374, 1.9406)	0.6839 (0.6832, 0.6846)
d'	0.8916 (0.8874, 0.8957)	0.8916 (0.8874, 0.8958)	1.0394 (1.0349, 1.0438)
	FFT (NY) with threshold	**FFT (NY) w/o threshold**	**FRS (EUT = 0.075) model**
Sensitivity	0.0196 (0.0195, 0.0196)	0.0152 (0.0152, 0.0153)	0.4348 (0.4346, 0.4349)
Specificity	0.9989 (0.9989, 0.9989)	0.9989 (0.9989, 0.9989)	0.8856 (0.8856, 0.8857)
LR+	18.4174 (5.4373, 62.3837)	14.3246 (4.1257, 49.7360)	3.8013 (3.3502, 4.3132)
LR–	0.9815 (0.9725, 0.9906)	0.9858 (0.9779, 0.9953)	0.6382 (0.6021, 0.6765)
c	2.5675 (2.5641, 2.5709)	2.6183 (2.6139, 2.6227)	0.6839 (0.6832, 0.6846)
d'	1.0094 (1.0031, 1.0157)	0.9078 (0.8997, 0.9160)	1.0394 (1.0349, 1.0438)
	FFT (NN) with threshold	**FFT (NN) w/o threshold**	**FRS (EUT = 0.075) model**
Sensitivity	0.0196 (0.0195, 0.0196)	0.0060 (0.0059, 0.0060)	0.4348 (0.4346, 0.4349)
Specificity	0.9989 (0.9989, 0.9989)	0.9998 (0.9998, 0.9998)	0.8856 (0.8856, 0.8857)
LR+	18.4174 (5.4373, 62.3837)	33.7405 (1.8674, 609.6319)	3.8013 (3.3502, 4.3132)
LR–	0.9815 (0.9725, 0.9906)	0.9942 (0.9892, 0.9992)	0.6382 (0.6021, 0.6765)
c	2.5675 (2.5641, 2.5709)	3.0430 (3.0349, 3.0511)	0.6839 (0.6832, 0.6846)
d'	1.0094 (1.0031, 1.0157)	1.0584 (1.0462, 1.0706)	1.0394 (1.0349, 1.0438)

*FFT only refers to two cues as the last cue has both exits. This FFT asks if the patient has diabetes (yes/no) (Y/N), is on blood pressure medication (BPRx) (Y/N) and is older than 44 (Y/N) (see Fig. 4).

[†]Assumes benefit/harms of statins for primary prevention of cardiovascular disease (CVD) of 2.5 and probability (pCVD) = 0.06; FFTT – fast and frugal tree that takes threshold in consideration; FFT – standard fast and frugal tree that takes no threshold into consideration (w/o); FRS – Framingham Risk Score.

[‡]ACC/AHA – The American College of Cardiology and American Heart Association recommends statins if pCVD ≥7.5% (see text for details).

[§]Note that because in both FFTT_ny and FFTT_nn trees 99.44% of the patients take the first exit, FFTT_ny and FFTT_nn result in almost identical results.

consequences of FFTT versus FFT versus FRS in terms of prescribing statins to people who actually end up developing CVD versus not prescribing statins to those who do not develop it.

Table 3 shows a comparison of treatments according to FFTT, FFT and AHA/ACC guidelines. In general, there is a statistically significant association – ranging from small to moderate – between FFT paths and risks for CVD. Note, however, the wide ranges in the probability estimates of CVD according to the FRS for each combination of the cues, which confirms the previously mentioned unreliability of the FRS estimates [31,32]. All models concur that patients who do not have diabetes, are not being treated for high blood pressure and are younger than 44 should not be prescribed statins ('NNN'). Likewise, all models agree that patients who have diabetes (irrespective of treatment for hypertension and age) should also be recommended statins. The models differ in their recommendations for patients with other risk profiles, but given the absence of universally agreed prediction and recommendations standards, it is not surprising that no clear-cut guideline for prescribing statins has emerged for all patients [23,26–28]. Indeed, the tendency to avoid unnecessary treatment becomes further apparent when the B/H ratio decreases: the *c* values become even more positive, indicating that our FFT/FFTT place more value on avoiding unnecessary treatment (i.e. false positives; results not shown). Of course, if our goal is to avoid false negatives, this FFT/FFTT performs only modestly, but so does FRS.

This consideration assumes that we weigh consequences of false negatives versus false positives equally. What would happen if a decision maker weighed these differently [33]? When the effect of differential weighting of false negatives (regret of omission) versus false positives (regret of commission) on the discriminatory capability of our FFT was assessed, avoidance of false positives was associated with both larger *d'* and larger *c* (Fig. 6). As Fig. 6 shows, under some circumstances our FFT (when weighted by expected consequences) can even achieve virtually perfect discrimination ($d' \geq 4$).

Table 3 Who should be treated with statins? Comparison of individualized decision making according to FFTT/evidence accumulation theory versus FFT versus FRS model/ACC/AHA guidelines

(a) FFTyy

FFTyy				Treatment		
Possible paths for this FFT†	pCVD* (median and range)	logLRsum**	Number of patients	According to FFTT*†	According to FTT*†	According to FRS model (ACC/AHA guidelines)*†
NNN	1.7% (0.43–24.5%)	–0.514	2732	No	No	No
NNY	7.45% (1.1–47.1%)	0.873	871	No	Yes	No
NY	11.9% (1.18–62.4%)	1.438	101	No	Yes	Yes
Y	21.7% (6.5–57.3%)	2.913	21	Yes	Yes	Yes
			3725			

*pCVD – probability of cardiovascular disease according to FRS (because highly skewed data, summary statistics expressed as median and range); FRS – Framingham Risk Score; ACC/AHA – The American College of Cardiology and American Heart Association recommends statins if pCVD ≥ 7.5%; FFTT – fast-and-frugal tree that takes threshold in consideration (according to the criterion stated in **); FFT – standard fast-and-frugal tree that takes no threshold into consideration but according to which each patient whose path ends with 'Y' should be treated (see text for details).

** $\sum_{a_i=1}\ln(LR_i+)+\sum_{a_i=0}\ln(LR_i-)$; treat if logLRsum $>ln\beta_{optimal}$ *(1.83; based on benefit/harms of statins of 2.5 and pCVD = 0.06)*; otherwise withhold treatment.

†Spearman rho correlation coefficient between pCVD and the FFT path: 0.64; $P < 0.000001$. Spearman rho correlation coefficient between pCVD and treatment according to: FFTT (0.11; $P < 0.000001$); FFT (0.60; $P < 0.000001$) and FRS/ACC/AHA guidelines (0.68; $P < 0.000001$).

(b) FFTyn

FFTyn				Treatment		
Possible paths for this FFT†	pCVD* (median and range)	logLRsum**	Number of patients	According to FFTT*†	According to FTT*†	According to FRS model (ACC/AHA guidelines)*†
NN	2.4% (0.43–47.1%)	–0.073	3603	No	No	No
NYN	6.4% (1.1–24.7%)	0.997	36	No	No	No
NYY	15.9% (2.5–62.4%)	2.38	65	Yes	Yes	Yes
Y	21.7% (6.5–57.3%)	2.913	21	Yes	Yes	Yes
			3725			

*pCVD – probability of cardiovascular disease according to FRS (because highly skewed data, summary statistics expressed as median and range); FRS – Framingham Risk Score; ACC/AHA – The American College of Cardiology and American Heart Association recommends statins if pCVD ≥ 7.5%; FFTT – fast-and-frugal tree that takes threshold in consideration (according to the criterion stated in **); FFT – standard fast-and-frugal tree that takes no threshold into consideration but according to which each patient whose path ends with 'Y' should be treated (see text for details).

** $\sum_{a_i=1}\ln(LR_i+)+\sum_{a_i=0}\ln(LR_i-)$; treat if logLRsum $>ln\beta_{optimal}$ *(1.83; based on benefit/harms of statins of 2.5 and pCVD = 0.06)*; otherwise withhold treatment.

†Spearman rho correlation coefficient between pCVD and the FFT path: 0.232; $P < 0.000001$. Spearman rho correlation coefficient between pCVD and treatment according to FFTT (0.223; $P < 0.000001$), FFT (0.223; $P < 0.000001$) and FRS/ACC/AHA guidelines (0.68; $P < 0.000001$).

(c) FFTny

FFTny				Treatment		
Possible paths for this FFT†	pCVD* (median and range)	logLRsum**	Number of patients	According to FFTT*†	According to FTT*†	According to FRS model (ACC/AHA guidelines)*†
N	2.5% (0.43–62.4%)	–0.018	3723	No	No	No
YNN	12.7% (8.9–15.9%)	2.41	4	Yes	No	Yes
YNY	24.9% (10.4–57.3%)	3.8	10	Yes	Yes	Yes
YY	35.4% (6.5–46.3%)	4.37	7	Yes	Yes	Yes
			3744			

*pCVD – probability of cardiovascular disease according to FRS (because highly skewed data, summary statistics expressed as median and range); FRS – Framingham Risk Score; ACC/AHA – The American College of Cardiology and American Heart Association recommends statins if pCVD ≥7.5%; FFTT – fast-and-frugal tree that takes threshold in consideration (according to the criterion stated in **); FFT – standard fast-and-frugal tree that takes no threshold into consideration but according to which each patient whose path ends with 'Y' should be treated (see text for details).

** $\sum_{a_i=1}\ln(LR_i+)+\sum_{a_i=0}\ln(LR_i-)$; treat if logLRsum $>ln\beta_{optimal}$ *(1.83; based on benefit/harms of statins of 2.5 and pCVD = 0.06)*; otherwise withhold treatment.

†Spearman rho correlation coefficient between pCVD and the FFT path: 0.115; $P < 0.000001$. Spearman rho correlation coefficient between pCVD and treatment according to FFTT (0.11; $P < 0.000001$), FFT (0.10; $P < 0.000001$) and FRS/ACC/AHA guidelines (0.68; $P < 0.000001$).

Table 3 *Continued*

(d) FFTnn

FFTnn				Treatment		
Possible paths for this FFT†	pCVD* (median and range)	logLRsum**	Number of patients	According to FFTT*†	According to FTT*†	According to FRS model (ACC/AHA guidelines)*†
N	2.5% (0.43–62.4%)	–0.018	3723	No	No	No
YN	18% (8.9–15.9%)	2.85	14	Yes	No	Yes
YYN	7.7% (6.5–8.8%)	3.92	2	Yes	No	Yes
YYY	35.5% (22.1–46.3%)	5.31	5	Yes	Yes	Yes
			3744			

*pCVD – probability of cardiovascular disease according to FRS (because highly skewed data, summary statistics expressed as median and range); FRS – Framingham Risk Score; ACC/AHA – The American College of Cardiology and American Heart Association recommends statins if pCVD ≥7.5%; FFTT – fast-and-frugal tree that takes threshold in consideration (according to the criterion stated in **); FFT – standard fast-and-frugal tree that takes no threshold into consideration but according to which each patient whose path ends with 'Y' should be treated (see text for details).

** $\sum_{a_i=1}\ln(LR_i+)+\sum_{a_i=0}\ln(LR_i-)$; treat if logLRsum $>ln\beta_{optimal}$ *(1.83; based on benefit/harms of statins of 2.5 and pCVD = 0.06)*; otherwise withhold treatment.

†Spearman rho correlation coefficient between pCVD and the FFT path: 0.115; $P < 0.000001$. Spearman rho correlation coefficient between pCVD and treatment according to FFTT (0.11; $P < 0.000001$), FFT (0.06; $P = 0.0008$) and FRS/ACC/AHA guidelines (0.68; $P < 0.000001$).

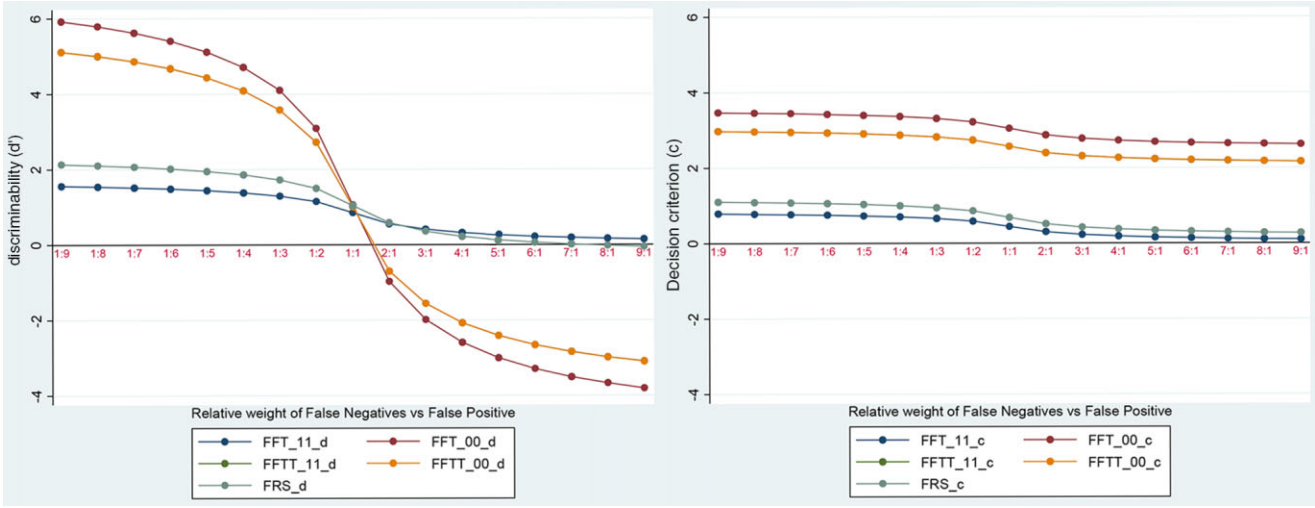

Figure 6 Effect of differential weighting of false negatives (regret of omission) versus false positives (regret of commission) on FFTs' discriminability (*d'*) and decision criterion (*c*) (for pCVD = 0.06 and B/H = 2.5). The effect on FFTyy, FFTTyy, FFTnn and FFTTnn is shown. Note how weighting may affect signal theory statistics to approach perfect discrimination under some circumstances ($d' > 4$).

Discussion

In this paper, we show how a simple threshold model [34] can function as a foundational building block that can serve as a link between SDT, FFTs and EAT. We draw attention to two issues of relevance for decision sciences. First, many seemingly different concepts used in different disciplines are actually equivalent. Second, by making this connection, we arrived at what we believe are novel findings. For one, Eq. (3) explicitly links the threshold model to EAT and FFT, thereby enabling creation of decision criteria that take into account both the accuracy statistics of FFT and the consequences built in the threshold model. As a result, decision recommendations based on standard FFT often differ from FFT that take threshold into consideration (FFTT), as demonstrated in this paper (see Table 3 and Fig. 5). Another novel result is that threshold criteria can be used as a strategy for selecting cues to construct an FFT (Table 1). Finally, although the threshold model describes a compensatory decision-making strategy and the FFT represents a non-compensatory approach to decision making, we show that both strategies are ultimately linked within a broader theoretical framework.

Of key interest, we believe that we described a new method for creating and evaluating FFTs of relevance to medical decision making. Clinical decision making is dominated by heuristic thinking and by simple decision trees resembling FFTs. Yet few methods have actually been developed to show how these FFTs can be constructed, evaluated and applied.

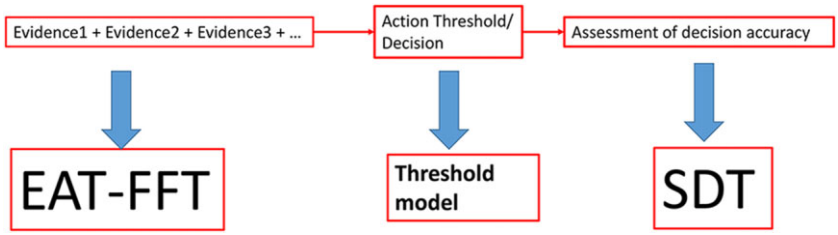

Figure 7 Model of decision making based on integration of evidenced accumulation theory (EAT), fast-and-frugal trees (FFT), threshold model and signal detection theory (SDT). FFT–EAT proposes that evidence is accumulated sequentially after which decision is made when certain threshold is exceeded (see Eq. 3). Such a decision is inevitably associated with false-positive versus false-negative errors, which are best appreciated within SDT framework (see Figs 1, 5 & 6).

Although the purpose of this paper is methodological, it is interesting to note that we succeeded in generating an FFT of practical importance. Consistent with FFT principles, we showed that our simple FFT, constructed of readily available cues, indicates that the patients who do not have diabetes and/or do not require antihypertensives can avoid statins. Likewise, we showed that patients who have diabetes should be treated. Both recommendations agree with the complex FRS model and AHA/ACC guidelines. In addition, it is instructive to note that our simple FFT departs from the AHA/ACA guidelines reflecting the disagreement in the field as to who should or should not be treated with statins [23,26–28,35]. This raises another interesting application of our proposed methodology: When an FFT does not agree with currently employed risk models and guideline recommendations (Table 3), this likely identifies open questions that need to be settled in future research.

In the final analysis, we showed how decisions depend on reliable information [36] – the results are highly sensitive to the estimate of the benefits and harms of interventions. Moreover, efficient cognitive processes require that we ignore part of the information available, particularly information that is unreliable, increases the estimation error or is costly to obtain [1,2]. Selection of a particular cue – the cue that may end up being used for constructing a simple FFT – should be evidence based and rely on as accurate information as possible. Specifically, we defined the criteria that can be used for selection of cues for FFT and under which circumstances such a selection would not be helpful.

In addition, we showed that the threshold model can provide a link between EUT and expected regret theory and SDT, FFT and EAT, thereby enabling us to integrate several theories of decision making under one conceptual umbrella [5]. By contrasting the logarithm of the sum of the accumulated evidence for positive (LR+) and negative (LR–) values for each cue with the action threshold, we define the decision criteria (for giving treatment). Most importantly, the latter offers an attractive possibility of individualizing treatment according to the combination of cues that a patient has.

Our method is not without limitations, however. Equation (3) assumes independence of cues, which may or may not be correct. Yet, as the bias–variance dilemma implies, assuming independence is likely to increase the bias of the model but simultaneously decrease estimation error (variance). Thus, the real question is one of the trade-off between the two sources of error: bias versus variance [2]. Future work should improve on the proposed methodology by taking into account this trade-off in the selection of cues.

Figure 7 shows a summary interpretation of our approach to unify several theoretical concepts. FFT–EAT proposes that evidence is accumulated sequentially after which decision is made when certain threshold is exceeded. Such a decision is inevitably associated with false-positive versus false-negative errors, which are best appreciated within SDT framework.

In conclusion, we showed that theories that at first glance appeared to be disconnected can be effectively linked under the theory integration programme [5]. Connecting apparently unrelated theories in different disciplines likely leads to discovery of new relationships. We believe that the programme of analysing the conceptual links between theoretical concepts can enable researchers from different disciplines to relate more fruitfully to each other.

Acknowledgement

This research is supported in part by DOD grant (#W81 XWH 09-2-0175) PI: Dr. Djulbegovic.

References

1. Gigerenzer, G. & Gaissmaier, W. (2011) Heuristic decision making. *Annual Review of Psychology*, 62, 451–482.
2. Gigerenzer, G. & Brighton, H. (2009) Homo heuristicus: why biased minds make better inferences. *Topics in Cognitive Science*, 1 (1), 107–143.
3. Luan, S., Schooler, L. J. & Gigerenzer, G. (2011) A signal-detection analysis of fast-and-frugal trees. *Psychological Review*, 118 (2), 316–338.
4. Lee, M. D. & Cummins, T. D. R. (2004) Evidence accumulation in decision making: unifying 'take the best' and 'rational' models. *Psychonomic Bulletin & Review*, 11 (2), 343–352.
5. Gigerenzer, G. (2015) A theory integration program. *Decision* (in press).
6. Pauker, S. G. & Kassirer, J. (1980) The threshold approach to clinical decision making. *The New England Journal of Medicine*, 302, 1109–1117.
7. Pauker, S. G. & Kassirer, J. P. (1975) Therapeutic decision making: a cost benefit analysis. *The New England Journal of Medicine*, 293, 229–234.
8. Mickes, L., Wixted, J. T. & Wais, P. E. (2007) A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review.*, 14 (5), 858–865.

9. Stanislaw, H. & Todorov, N. (1999) Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31 (1), 137–149.
10. Gigerenzer, G., Hertwig, R. & Pachur, T. (eds) (2011) Heuristics. The Foundation of Adaptive Behavior. New York: Oxford University Press.
11. Jenny, M. A., Pachur, T., Williams, S. L., Becker, E. & Margraf, J. (2013) Simple rules for detecting depression. *Journal of Applied Research in Memory and Cognition*, 2, 149–157.
12. Hozo, I. & Djulbegovic, B. (2008) When is diagnostic testing inappropriate or irrational? Acceptable regret approach. *Medical Decision Making*, 28 (4), 540–553.
13. Hozo, I. & Djulbegovic, B. (2009) Clarification and corrections of acceptable regret model. *Medical Decision Making*, 29, 323–324.
14. Tsalatsanis, A., Hozo, I., Vickers, A. & Djulbegovic, B. (2010) A regret theory approach to decision curve analysis: a novel method for eliciting decision makers' preferences and decision-making. *BMC Medical Informatics and Decision Making*, 10 (1), 51.
15. Djulbegovic, B., Elqayam, S., Reljic, T., *et al*. (2014) How do physicians decide to treat: an empirical evaluation of the threshold model. *BMC Medical Informatics and Decision Making*, 14 (1), 47.
16. Martignon, L., Katsikopoulos, K. V. & Woikeb, J. K. (2008) Categorization with limited resources: a family of simple heuristics. *Journal of Mathematical Psychology*, 52, 352–361.
17. Sox, H. C., Higgins, M. C. & Owens, D. (2013) Medical Decision Making, 2nd edn. Chicester: Wiley-Blackwell.
18. Cholesterol Treatment Trialists C (2012) The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *The Lancet*, 380 (9841), 581–590.
19. Taylor, F., Huffman, M. D., Macedo, A. F., *et al*. (2013) Statins for the primary prevention of cardiovascular disease. *The Cochrane Database of Systematic Reviews*, (1), CD004816.
20. Stone, N. J., Robinson, J. G., Lichtenstein, A. H., *et al*. (2014) 2013 ACC/AHA Guideline on the Treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *Journal of the American College of Cardiology*, 63 (25 Pt B), 2889–2934.
21. D'Agostino, R. B. Sr, Vasan, R. S., Pencina, M. J., *et al*. (2008) General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*, 117 (6), 743–753.
22. Framingham Heart Study. (2015) Framingham Heart Study. Available at: https://www.framinghamheartstudy.org/.
23. Pencina, M. J., Navar-Boggan, A. M., D'Agostino, R. B. Sr, *et al*. (2014) Application of new cholesterol guidelines to a population-based sample. *The New England Journal of Medicine*, 370 (15), 1422–1431.
24. Newman, D. (2015) Statin drugs given for 5 years for heart disease prevention (without known heart disease). Available at: http://www.thennt.com/nnt/statins-for-heart-disease-prevention-without-prior-heart-disease/ (last accessed 4 September 2015).
25. Newman, C. B. & Tobert, J. A. (2015) Statin intolerance: reconciling clinical trials and clinical experience. *JAMA: The Journal of the American Medical Association*, 313 (10), 1011–1012.
26. Redberg, R. F. & Katz, M. H. (2012) Healthy men should not take statins. *JAMA: The Journal of the American Medical Association*, 307 (14), 1491–1492.
27. Ioannidis, J. P. (2014) Guidelines for cardiovascular risk assessment and cholesterol treatment – reply. *JAMA: The Journal of the American Medical Association*, 311 (21), 2235–2236.
28. Ioannidis, J. P. (2014) More than a billion people taking statins? Potential implications of the new cardiovascular guidelines. *JAMA: The Journal of the American Medical Association*, 311 (5), 463–464.
29. Center for Disease and Control (2010) Prevalence of Coronary Heart Disease – United States, 2006–2010.
30. Wickens, T. D. (2002) Elementary Signal Detection Theory. Oxford: Oxford University Press.
31. DeFilippis, A. P., Young, R., Carrubba, C. J., *et al*. (2015) An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort: calibration and discrimination among CVD risk scores. *Annals of Internal Medicine*, 162 (4), 266–275.
32. Ridker, P. M. & Cook, N. R. (2015) Comparing cardiovascular risk prediction scores. *Annals of Internal Medicine*, 162 (4), 313–314.
33. Zeelenberg, M. & Pieters, R. (2007) A theory of regret regulation 1.0. *Journal of Consumer Psychology: The Official Journal of the Society for Consumer Psychology*, 17, 3–18.
34. Djulbegovic, B., van den Ende, J., Hamm, R. M., *et al*. (2015) When is rational to order a diagnostic test, or prescribe treatment: the threshold model as an explanation of practice variation. *European Journal of Clinical Investigation*, 45 (5), 485–493.
35. Montori, V. M., Brito, J. P. & Ting, H. H. (2014) Patient-centered and practical application of new high cholesterol guidelines to prevent cardiovascular disease. *JAMA: The Journal of the American Medical Association*, 311 (5), 465–466.
36. Djulbegovic, B., Guyatt, G. H. & Ashcroft, R. E. (2009) Epistemologic inquiries in evidence-based medicine. *Cancer Control: Journal of the Moffitt Cancer Center*, 16 (2), 158–168.
37. Djulbegovic, B. & Hozo, I. (2007) When should potentially false research findings be considered acceptable? *PLoS Medicine*, 4 (2), e26.
38. Hozo, I. & Djulbegovic, B. (2009) Will insistence on practicing medicine according to expected utility theory lead to an increase in diagnostic testing? *Medical Decision Making*, 29, 320–322.

Appendix A

Signal detection theory terminology*

Term	Definition/measurement	Comments
Hit rate (H), or sensitivity (S)	Respond 'signal' given signal	Miss rate or false-negative rate (FN) = 1 – S
False alarm rate (FA) or false-positive rate (FP)	Respond 'signal' given noise	FA = 1 – C (specificity) (or, correct rejection) (CR)
d' (discriminability)	*z*Hit – *z*FA	*z*Hit and *z*FA are the *z*-scores of the hit rate and the false alarm rate, respectively; *d'* reflects the standardized distance between the signal (S) and noise (N) distributions
c (decision criterion)	$-0.5 \times (z\text{Hit} - z\text{FA})$	$c > 0$: *conservative* bias, making more noise than signal decisions relative to prior probabilities $c < 0$: *liberal* bias, making more signal than noise decisions relative to prior probabilities $c = 0$: *neutral* bias, making decisions consistent with prior probabilities
$\beta_{optimal}$	$\beta_{optimal} = \frac{V(C) - V(FA)}{V(S) - V(FN)} \cdot \frac{p(Noise)}{p(Signal)}$	*P*(N) and *P*(S) are prior probabilities of the two decision events; *V*(Hit), *V*(miss), *V*(CR) and *V*(FA) are expected values of the four decision outcomes; with a certain *d'*: $c_{optimal} = \ln(\beta_{optimal})/d$.

*From Luan *et al.* [3] (note that we use terminologies common in psychological and in medical fields).

Appendix B

A brief derivation of the threshold model (see details in 6,7,12,13,37)

Using the decision tree represented in Fig. 2, we have the following expected values:

Expected utility	or	Expected regret
$EU[Rx] = p \cdot U_1 + (1 - p) \cdot U_2$		$ER[Rx] = (1 - p)(U_4 - U_2)$
$EU[NoRx] = p \cdot U_3 + (1 - p) \cdot U_4$		$ER[NoRx] = p(U_1 - U_3)$

To answer the question 'With which probability of the disease, $p = pD$, will the expected utility (or regret) value of treating the patient be larger (smaller) than the expected value of not treating?', we need to solve the inequality:

Expected utility	or	Expected regret
$EU[NoRx] \le EU[Rx]$		$ER[NoRx] \ge ER[Rx]$
$pU_3 + (1 - p)U_4 \le pU_1 + (1 - p)U_2$		$p(U_1 - U_3) \ge (1 - p)(U_4 - U_2)$

Solving either of these for the probability, we get the inequality:

$$U_4 - U_2 \le p \cdot U_1 - p \cdot U_3 + p \cdot U_4 - p \cdot U_2$$

$$p \ge \frac{U_4 - U_2}{U_1 - U_3 + U_4 - U_2}$$

If we define the net benefits of the treatment Rx, as $B = U_1 - U_3$ and the net harms of the treatment as $H = U_4 - U_2$, we can rewrite the formula earlier by defining the above right-hand expression as the threshold probability:

$$T = \frac{H}{B + H} = \frac{1}{1 + \frac{B}{H}}$$

In conclusion, when faced with a decision whether to prescribe treatment or not, we simply compare the probability, $p = pD$, with the threshold and

1 Prescribe treatment if $p \ge T = \frac{1}{1 + \frac{B}{H}}$

2 Withhold treatment if $p < T = \frac{1}{1 + \frac{B}{H}}$

Appendix C

Expected utility theory (threshold model)

$p = p(D+) =$ the prior probability of disease

$1 - p = p(D-) =$ the prior probability of no disease

Test (T) has two possible outcomes: positive test ($T+$) and negative test ($T-$).

$P(T+|D+) =$ True positive $= S = Sensitivity$

$P(T-|D-) =$ True negative $= C = S_p = Specificity$

$P(T-|D+) =$ False negative $= 1 - S$

$P(T+|D-) =$ False positive $=$ False alarm $= 1 - C$

With utilities
U_1: Utility of true positive
U_2: Utility of false positive
U_3: Utility of false negative
U_4: Utility of true negative

The benefits and harms are defined as

$$B = U_1 - U_3 \quad and \quad H = U_4 - U_2$$

Post-test probability value of correctly identifying the disease is given as $PPV = P(T+|D+)$:

$$PPV = \frac{P(T+|D+)P(D+)}{P(T+|D+)P(D+)+P(T+|D-)P(D-)} = \frac{S\cdot p}{S\cdot p+(1-C)(1-p)}$$

The likelihood ratio (positive) of the test is given as

$$LR+ = \frac{S}{1-C} = \frac{P(T+|D+)}{P(T+|D-)}$$

According to expected utility theory, treatment is prescribed if

$$PPV = P(D+|T+) \geq \frac{1}{1+\frac{B}{H}}$$

or in other words, after a bit of algebra, when

$$\frac{1-PPV}{PPV} \leq \frac{B}{H}$$

The same expression can be expressed using the odds ratio of post-test probabilities and likelihood ratios:

$$\frac{1-PPV}{PPV} = \frac{(1-C)(1-p)}{S\cdot p} = \frac{1-C}{S}\cdot\frac{1-p}{p}$$

Therefore, treatment will be prescribed whenever

$$\frac{1}{LR+}\cdot\frac{1-p}{p} \leq \frac{B}{H} \quad or \quad LR+ \geq \frac{H}{B}\cdot\frac{1-p}{p}$$

Signal detection theory

$p = p(Signal) =$ prior probability of signal

$1-p = p(Noise) =$ prior probability of noise

One needs to respond to a stimulus and decide whether it is 'signal' or 'noise'. So $T+ = Signal$, $T- = Noise$, $D+ = Signal$, $D- = Noise$:

$P(T+|D+) =$ True positive (hit)

$P(T-|D-) =$ True negative (correct rejection = CR)

$P(T-|D+) =$ False negative (miss)

$P(T+|D-) =$ False positive (false alarm = FA)

With utilities
$V(Hit)$: Utility of hit (true positive)
$V(FA)$: Utility of false alarm (false positive)
$V(Miss)$: Utility of miss (false negative)
$V(CR)$: Utility of correct rejection (true negative)

The benefits and harms could be defined as

$$B = V(Hit)-V(Miss) \quad and \quad H = V(CR)-V(FA)$$

Post-test probability value of correctly identifying the signal is given as

$$PPV = \frac{P(T+|D+)P(D+)}{P(T+|D+)P(D+)+P(T+|D-)P(D-)} = \frac{P(Hit)P(Signal)}{P(Hit)P(Signal)+P(FA)P(Noise)}$$

The likelihood ratio is given as (f = density distribution)

$$LR+ = \frac{P(T+|Signal)}{P(T+|Noise)} = \frac{P(Hit)}{1-P(CR)}$$

Then, the decision is made in favor of signal by comparing the likelihood ratio with $\beta_{optimal}$:

$$LR+ \geq \beta_{optimal} = \frac{V(CR)-V(FA)}{V(Hit)-V(Miss)}\cdot\frac{P(Noise)}{P(Signal)}$$

or in other words, when

$$LR+ \geq \frac{H}{B}\cdot\frac{1-p}{p}$$

$\beta_{optimal}$ is the equivalent of the threshold for likelihood ratio(s). In other words, comparing probability with threshold (P_t) is equivalent to comparing likelihood ratio(s) with $\beta_{optimal}$.

Therefore,

If $(LR-) \geq \beta_{optimal}$, then even if test/cue proves to be negative/noise,

$$(LR-) \geq \frac{H}{B}\cdot\frac{1-p}{p} \Rightarrow \frac{1}{(LR-)}\cdot\frac{1-p}{p} \leq \frac{B}{H} \Rightarrow \frac{C}{1-S}\cdot\frac{1-p}{p} \leq \frac{B}{H} \Rightarrow \frac{1-P(D+|T-)}{P(D+|T-)} \leq \frac{B}{H} \Rightarrow P(D+|T-) \geq \frac{1}{1+\frac{B}{H}}$$

Hence, when $(LR-) \geq \beta_{optima}$, the post-test probability is *always larger* than the threshold probability:

$$\frac{1}{1+\frac{B}{H}} \leq P(D+|T-) \leq P(D+|T+)$$

If $(LR+) \leq \beta_{optimal}$, then even if test/cue proves to be positive/signal,

$$(LR+) \leq \frac{H}{B}\cdot\frac{1-p}{p} \Rightarrow \frac{1}{(LR+)}\cdot\frac{1-p}{p} \geq \frac{B}{H} \Rightarrow \frac{1-C}{S}\cdot\frac{1-p}{p} \geq \frac{B}{H} \Rightarrow \frac{1-P(D+|T+)}{P(D+|T+)} \geq \frac{B}{H} \Rightarrow P(D+|T+) \leq \frac{1}{1+\frac{B}{H}}$$

Hence, when $(LR+) \leq \beta_{optimal}$, the post-test probability is *always smaller* than threshold probability:

$$P(D+|T-) \leq P(D+|T+) \leq \frac{1}{1+\frac{B}{H}}$$

Thus, $\beta_{optimal}$ is equivalent to the action threshold in the threshold model.

Appendix D

Linking the threshold model to evidence accumulation theory

We consider a typical but specific question: Is a *signal* present? As explained in the manuscript, the absence of a signal represents *noise*. Alternatively, in clinical terms, we have a single question: Does the patient have a Disease (D+)? Absence of the disease is represented by No Disease (D–). Let us define $p = p(D+)$ as the prior probability of the disease (prevalence).

Each cue C_i, where $i = 1, 2 \ldots k$ has a binary variable: $a_i = 1$(positive) or $a_i = 0$(negative), measuring a particular symptom, test result, observation . . .

Hence, each patient has a sequence of cue answers: (a_1, a_2, . . . , a_k).

Define, for a particular cue C_i as:

$$\text{True positive } TP_i = p(a_i = 1 \cap D+)$$

and True positive rate (S_i = sensitivity)

$$TPR_i = p(a_i = 1|D+) = \frac{p(a_i = 1 \cap D+)}{p(D+)}$$

$$\text{True negative } TN_i = p(a_i = 0 \cap D-)$$

and True negative rate (C_i = specificity)

$$TNR_i = p(a_i = 0|D-) = \frac{p(a_i = 0 \cap D-)}{p(D-)}$$

Using Bayes' theorem results in

$$\text{Positive-predictive value } PPV_i = p(D+|a_i = 1)$$
$$= \frac{p(a_i = 1|D+)\,p(D+)}{p(a_i = 1|D+)\,p(D+) + p(a_i = 1|D-)\,p(D-)}$$
$$= \frac{TPR_i \cdot p}{TPR_i \cdot p + (1 - TNR_i) \cdot (1 - p)} = \frac{S_i \cdot p}{S_i \cdot p + (1 - C_i) \cdot (1 - p)}$$

$$\text{Negative-predictive value } NPV_i = p(D-|a_i = 0)$$
$$= \frac{p(a_i = 0|D-)\,p(D-)}{p(a_i = 0|D+)\,p(D+) + p(a_i = 0|D-)\,p(D-)}$$
$$= \frac{TNR_i \cdot (1 - p)}{(1 - TPR_i) \cdot p + TNR_i \cdot (1 - p)} = \frac{C_i \cdot (1 - p)}{(1 - S_i) \cdot p + C_i \cdot (1 - p)}$$

Let us write these using odds ratio and likelihood ratio. Recall that

$$(LR+) = \frac{S}{1 - C} \quad \text{and} \quad (LR-) = \frac{1 - S}{C}$$

Therefore, the posterior odds for the presence of the disease (whether the cue is positive or negative) are

$$\frac{P(D+|T+)}{1 - P(D+|T+)} = \frac{PPV_i}{1 - PPV_i} = \frac{S_i \cdot p}{(1 - C_i) \cdot (1 - p)} = (LR_i+) \cdot \frac{p}{1 - p}$$

$$\frac{P(D+|T-)}{1 - P(D+|T-)} = \frac{1 - NPV_i}{NPV_i} = \frac{(1 - S_i) \cdot p}{C_i \cdot (1 - p)} = (LR_i-) \cdot \frac{p}{1 - p}$$

According to Martignon [16], we define two validities

$v_i^1 = p(D+|a_i = 1) = PPV_i$(*positive-predictive value*) and

$v_i^0 = p(D-|a_i = 0) = NPV_i$(*negative-predictive value*)

Then, we define the overall validity/accuracy of the cue as

$$v_i^p = \frac{correct}{all\ decisions} = \frac{TP + TN}{TP + FP + TN + FN}$$
$$= p(a_i = 1 \cap D+) + p(a_i = 0 \cap D-)$$
$$= p(a_i = 1|D+)\,p(D+) + p(a_i = 0 \cap D-)\,p(D-)$$
$$= TPR_i \cdot p + TNR_i \cdot (1 - p)$$

The rule of **MAX** states the following: 'Order cues by decreasing the value of $\max\{v_i^1, v_i^0\} = \max\{PPV_i, NPV_i\}$'. The rule of MAX also states that if $PPV_i > NPV_i$, or in other words $v_i^1 = \max\{v_i^1, v_i^0\}$, then the exit is on the left ($a_i = 1$); otherwise, the exit is on the right ($a_i = 0$).

The rule of **ZIG** states the following: First, compute the ratio $r = \max\left\{\frac{p(D+)}{p(D-)}, \frac{p(D-)}{p(D+)}\right\} = \max\left\{\frac{p}{1-p}, \frac{1-p}{p}\right\}$. That means that $1 \le r$. Then find an integer k, such that $2^{k-1} \le r < 2^k$. In other words, if $p(D+) = p(D-)$, then $p = 1 - p$, meaning that $r = 1$, and $k = 1$ because $2^{1-1} \le 1 < 2^1$. If, on the other hand, the disease is extremely unlikely, $p(D+) = 0.01$, then $\frac{1-p}{p} = \frac{0.99}{0.01} = 99$, meaning that $k = 7$ because $64 = 2^{7-1} \le 99 < 2^7 = 128$.

Once we have the number k, we assign *all levels from 1 to k* to be the same: left ($a_i = 1$) if $p > 1 - p$ and right ($a_i = 0$) if $p < 1 - p$. The rest of the exits will *alternate*. Now that the pattern of exits is known (zigzag), we determine the *cue order*: Start with level 1: if the exit is on the left ($a_i = 1$), pick the cue with the highest $v_i^1 = PPV_i$; if the exit is on the right ($a_i = 0$), pick the cue with the highest $v_i^0 = NPV_i$. Repeat, level by level, until the $(n - 1)^{st}$ level.

Evidence accumulation theory

Following Lee and Cummins' definition of the rational model (RAT) [4], if we have k cues, the log odds ratio is then defined as

$$L_D = \ln\frac{p(D+|a_1, a_2, \ldots, a_k)}{p(D-|a_1, a_2, \ldots, a_k)} = (\textit{using Bayes's theorem})$$
$$= \ln\frac{\frac{p(a_1, a_2, \ldots, a_k|D+)\,p(D+)}{p(a_1, a_2, \ldots, a_k)}}{\frac{p(a_1, a_2, \ldots, a_k|D-)\,p(D-)}{p(a_1, a_2, \ldots, a_k)}}$$
$$= \ln\frac{p(a_1, a_2, \ldots, a_k|D+)\,p(D+)}{p(a_1, a_2, \ldots, a_k|D-)\,p(D-)}$$
$$= \ln\frac{p(a_1, a_2, \ldots, a_k|D+)}{p(a_1, a_2, \ldots, a_k|D-)} + \ln\frac{p(D+)}{p(D-)}$$

Assuming independence of the cues, $p(a_1, a_2, \ldots, a_k|D+) = \prod_i p(a_i|D+)$, and assuming that $p = p(D+)$ is the prior probability of disease (prevalence):

$$L_D = \sum_{i=1}^{k} \ln\frac{p(a_i|D+)}{p(a_i|D-)} + \ln\frac{p}{1-p}$$

For each patient and for each cue, C_i, there are two possibilities:

1 $a_i = 1$: In this case, again using Bayes' theorem results in

$$\ln\frac{p(a_i = 1|D+)}{p(a_i = 1|D-)} = \ln\frac{\frac{p(D+\,|\,a_i = 1)\,p(a_i = 1)}{p(D+)}}{\frac{p(D-|a_i = 1)\,p(a_i = 1)}{p(D-)}}$$
$$= \ln\frac{p(D+|a_i = 1)}{p(D-|a_i = 1)} - \ln\frac{p(D+)}{p(D-)}$$
$$= \ln\frac{PPV_i}{1 - PPV_i} - \ln\frac{p}{1-p}$$

2 $a_i = 0$: In this case, again using Bayes' theorem results in

$$\ln\frac{p(a_i=0|D+)}{p(a_i=0|D-)}=\ln\frac{\frac{p(D+|a_i=0)p(a_i=0)}{p(D+)}}{\frac{p(D-|a_i=0)p(a_i=0)}{p(D-)}}$$
$$=\ln\frac{p(D+|a_i=0)}{p(D-|a_i=0)}-\ln\frac{p(D+)}{p(D-)}$$
$$=\ln\frac{1-NPV_i}{NPV_i}-\ln\frac{p}{1-p}$$

Therefore, the final formula for each patient is:

$$L_D=\sum_{a_i=1}\left(\ln\frac{PPV_i}{1-PPV_i}-\ln\frac{p}{1-p}\right)+\sum_{a_i=0}\left(\ln\frac{1-NPV_i}{NPV_i}-\ln\frac{p}{1-p}\right)+\ln\frac{p}{1-p}$$

According to RAT, if $L_D>0$, the patient is assumed to have the disease, and if $L_D<0$, the patient does not have the disease.

Recall from our threshold model (Appendices B and C):

$$\frac{PPV_i}{1-PPV_i}=(LR_i+)\cdot\frac{p}{1-p}\quad\text{so}\quad\ln\frac{PPV_i}{1-PPV_i}=\ln(LR_i+)+\ln\frac{p}{1-p}$$

$$\frac{1-NPV_i}{NPV_i}=(LR_i-)\cdot\frac{p}{1-p}\quad\text{so}\quad\ln\frac{1-NPV_i}{NPV_i}=\ln(LR_i-)+\ln\frac{p}{1-p}$$

Combining these expressions, we get

$$L_D=\sum_{a_i=1}\ln(LR_i+)+\sum_{a_i=0}\ln(LR_i-)+\ln\frac{p}{1-p}$$

Note that L_D = posterior odds for all cues (the entire FFT tree). In other words, for each patient with a sequence of cue answers (a_1, a_2, . . . , a_k), the log odds ratio earlier can be evaluated to decide whether the patient has the disease or not.

Threshold model

As previously discussed, according to expected utility theory (or expected regret theory) [13,37,38], treatment should be prescribed when the posterior probability after all the cues (*PP*):

$$\frac{PP}{1-PP}=\frac{1-p(D+|a_1,a_2,\ldots,a_k)}{p(D+|a_1,a_2,\ldots,a_k)}\geq\frac{B}{H}\quad or$$
$$\frac{p(D+|a_1,a_2,\ldots,a_k)}{p(D-|a_1,a_2,\ldots,a_k)}\leq\frac{H}{B}$$

Taking the logarithm of both sides results in

$$\ln\frac{p(D+|a_1,a_2,\ldots,a_k)}{p(D-|a_1,a_2,\ldots,a_k)}\leq\ln\frac{H}{B}$$

Or, using the formula for the L_D above:

$$L_D=\sum_{a_i=1}\ln(LR_i+)+\sum_{a_i=0}\ln(LR_i-)+\ln\frac{p}{1-p}\leq\ln\frac{H}{B}$$

Therefore, the refinement of the RAT model with utilities/regret can be expressed as

$$\sum_{a_i=1}\ln(LR_i+)+\sum_{a_i=0}\ln(LR_i-)\leq\ln\frac{H}{B}-\ln\frac{p}{1-p}=\ln\frac{H}{B}\cdot\frac{1-p}{p}$$
$$=\ln\beta_{optimal}$$

Or:

$$\sum_{a_i=1}\ln(LR_i+)+\sum_{a_i=0}\ln(LR_i-)\leq\ln\beta_{optimal}$$

In other words, for each patient with a sequence of cue answers(a_1, a_2, . . . , a_k), the sum of the logs of the likelihood ratios earlier can be evaluated to see whether it is smaller than the log of $\beta_{optimal}$. If the left-hand sum is smaller than $\ln\beta_{optimal}$, treatment should be withheld and if it is larger, treatment should be prescribed.

Appendix E

Discrimination or discriminability index *d'* (the distance between the signal and the noise means in standard deviation units)

For a given model, we can calculate this measure using true positives (TP, hits), false positives (i.e. false alarms, FA), false negatives (FN, misses) and true negatives (TN, correct rejections).

$$d'=\Phi^{-1}(TP)-\Phi^{-1}(FA),$$

where Φ^{-1} is the inverse standard normal function. In terms of *z*-scores, $zH=\Phi^{-1}(TP)=\Phi^{-1}(Hit)$ and $zF=\Phi^{-1}(FA)=\Phi^{-1}(false\ alarms)$, we often rewrite the formula above as:

$$d'=zH-zF$$

To weigh the FN (=1 − TP) and FA differently, the weight parameter *w* is used so that:

$$d'=w\cdot zH-(2-w)\cdot zF$$

For example, if the weight ratio of FN to FA is 1:3 (in other words, if false positives are considered three times more important than false negatives), we need to solve the equation:

$$w:(2-w)=1:3\quad\text{or}\quad\frac{w}{2-w}=\frac{1}{3},\quad\text{which has solution w}=0.5$$

We calculated weighted *d'* for a variety of proportions.

The decision criterion *c* (midpoint between signal and the noise means in standard deviation units)

The decision criterion is defined (just like *d'*) through the *z*-scores:

$$c=-\frac{zH+zF}{2}$$

To weigh the TP and FP differently, we simply use the weight parameter *w*, so that

$$c=-\frac{w\cdot zH+(2-w)\cdot zF}{2}$$

Here the same values of the weight *w* can be used as above in calculations of *d'*.

Variances of discriminability indices and/or decision criteria (and differences)

A standard procedure to test the difference between two independent measures is usually using the *z*-test, where

$$z = \frac{x_1 - x_2}{\sqrt{var(x_1) + var(x_2)}}$$

Using this idea, for different models, we can estimate the test statistic. In the case of decision criteria or discriminability index, we have

$$z_c = \frac{c_1 - c_2}{\sqrt{var(c_1) + var(c_2)}} \quad \text{and} \quad z_{d'} = \frac{d_1' - d_2'}{\sqrt{var(d_1') + var(d_2')}}$$

According to Wickens [30] the variances of d_i' and the decision criterion can be estimated as

$$var(c_i) = \frac{var(F_i)}{\varphi^2(c_i)} = \frac{\frac{F_i \cdot (1-F_i)}{N_n}}{\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{c_i^2}{2}}\right)^2} = \frac{\frac{F_i \cdot (1-F_i)}{N_n}}{\frac{1}{2\pi} e^{-c_i^2}}$$

$$= \frac{2\pi e^{c_i^2}}{N_n} \cdot F_i \cdot (1-F_i) \quad \text{and}$$

$$var(d_i') = \frac{var(F_i)}{\varphi^2(c_i)} + \frac{var(H_i)}{\varphi^2(d_i' - c_i)}$$

$$= \frac{2\pi e^{c_i^2}}{N_n} \cdot F_i \cdot (1-F_i) + \frac{2\pi e^{(d_i' - c_i)^2}}{N_s} \cdot H_i \cdot (1-H_i)$$

Here $F_i = false\ alarm\ rate = 1 - specificity$, $H_i = hit\ rate = sensitivity$, and N_n and N_s are the sample size of the noise and signal, respectively. The function φ is the standard normal distribution density $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

In the case of d_i', we can simplify the formula by assuming that the two measures have the same noise distribution, such that

$$d_1' - d_2' = (zH_1 - zF) - (zH_2 - zF) = zH_1 - zH_2,$$

which makes the variance

$$var(d_1' - d_2') = \frac{2\pi e^{(d_1' - c_1)^2}}{N_s} \cdot H_1 \cdot (1-H_1) + \frac{2\pi e^{(d_2' - c_2)^2}}{N_s} \cdot H_2 \cdot (1-H_2)$$

In summary, we can use the variances above for 95% confidence intervals of *d'* and *c*, and according to Wickens, for 'moderate *d'* and large samples' [30].

The test compares the value of $|z_c|$ or $|z_{d'}|$ with $z_{\alpha/2}$ (which is 1.96 for $\alpha = 5\%$). If $|z| > z_{\alpha/2}$ we will reject the hypothesis that $d_1' = d_2'$; if $|z| \le z_{\alpha/2}$, we will not be able to reject the hypothesis that $d_1' = d_2'$.

Likelihood ratio

If $LR = \frac{a/b}{c/d}$ (a ratio of two proportions), the usual method for finding the variance is to use logarithmic transformation, which is approximately normally distributed with variance $\sigma^2 = \frac{1}{a} - \frac{1}{b} + \frac{1}{c} - \frac{1}{d}$ and the confidence interval [$LR \cdot e^{-1.96 \cdot \sigma}$, $LR \cdot e^{1.96 \cdot \sigma}$].