Treatment Threshold

Diagnosis Threshold

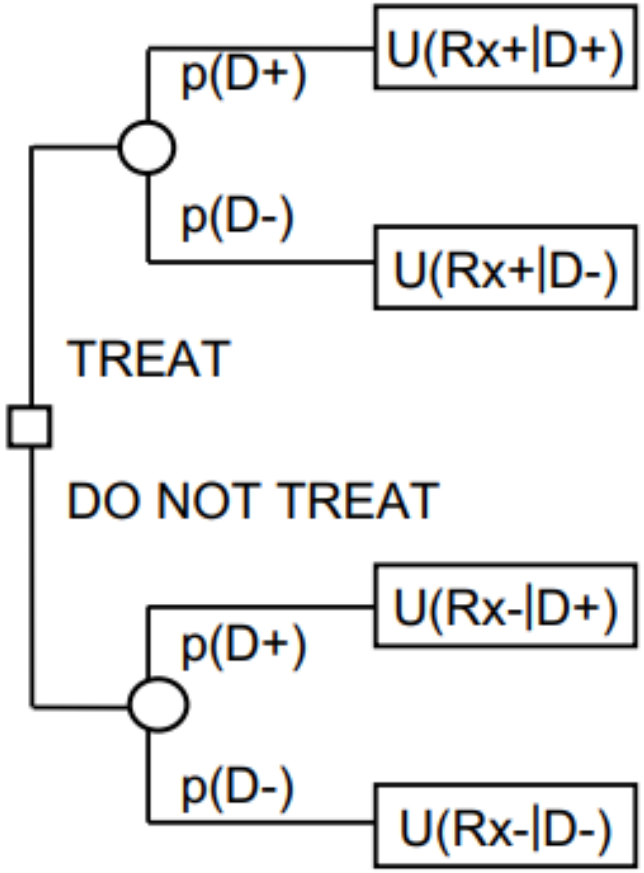
Ali Kabir, MD, MPH, MSc, PhD

Associate Professor of Epidemiology, Resident of Internal Medicine

Firoozgar Hospital, Iran University of Medical Sciences

Clinical practice: how decisions are made

To treat or not to treat decision tree



Expected Utility of treatment:

EU(Rx+) = p(D+) x U(Rx+|D+) + p(D-) x U(Rx+|D-)

Expected Utility of no treatment:

EU(Rx-) = p(D+) x U(Rx-|D+) + p(D-) x U(Rx-|D-)

Harm of Rx = U(Rx-|D-) ***minus*** U(Rx+|D-)

Benefit of Rx = U(Rx+|D+) ***minus*** U(Rx-|D+)

Harm and benefit of Rx

Harm of Rx = U(Rx-|D-) ***minus*** U(Rx+|D-)
= ***loss*** in health when untreated patients without disease are compared with treated patients without disease
(note that the value must be positive as we will always have U(Rx-|D-) > U(Rx+|D-)!)

Benefit of Rx = U(Rx+|D+) ***minus*** U(Rx-|D+)
= ***net gain*** in health when treated patients with disease are compared with untreated patients with disease
Note: ***net*** because side-effects of Rx must be taken into account!

Example: deep tumour in the brain of a 40-year old male, high-risk operation (operative mortality 2%), definitive diagnosis can only be made during operation. D+ = malignant; D- = benign
Assume life expectancy (LE) if we don't operate is about 40 years (can be found in LE tables)

Example of utilities:

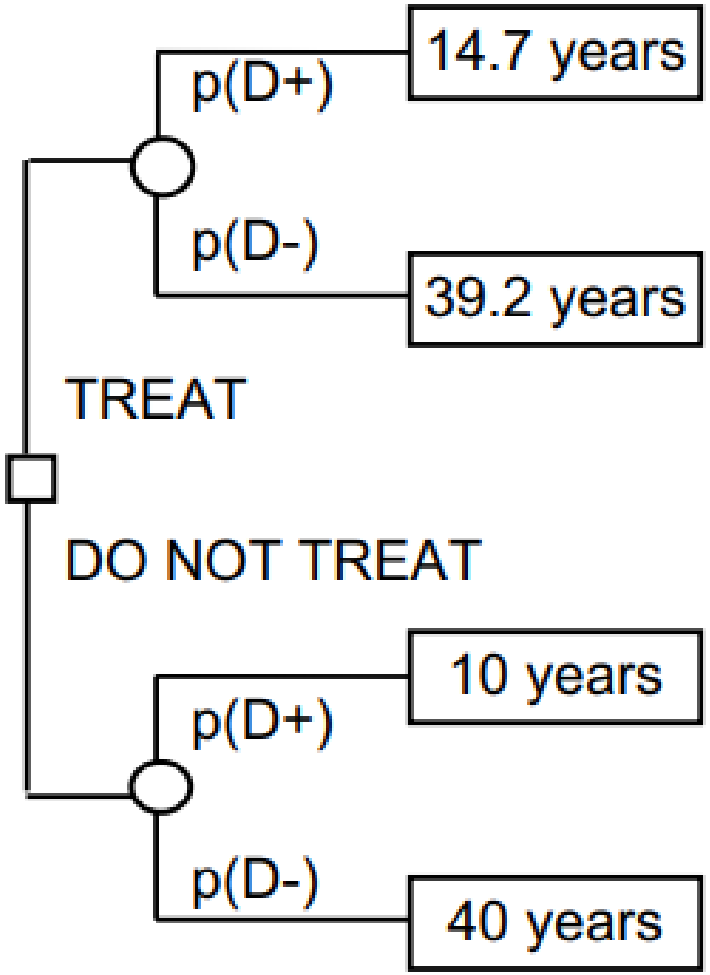
U of operating a malignant tumour: U(Rx+|D+)
Assume LE = 15 years ***in survivors*** of the operation
LE = 0.02 x 0 + 0.98 x 15 = 14.7
= prob. of death x LE|death + prob. survive x LE|survive

U of operating a benign tumour: U(Rx+|D-)
LE is 0.98 x 40 = 39.2 years

U of not operating a malignant tumour:
U(Rx-|D+)
Assume LE = 10 years

U of not operating a benign tumour: U(Rx-|D-)
LE = 40 years

The tree with these utilities



Expected Utility of treatment:
EU(Rx+) = p(D+) x 14.7 + p(D-) x 39.2

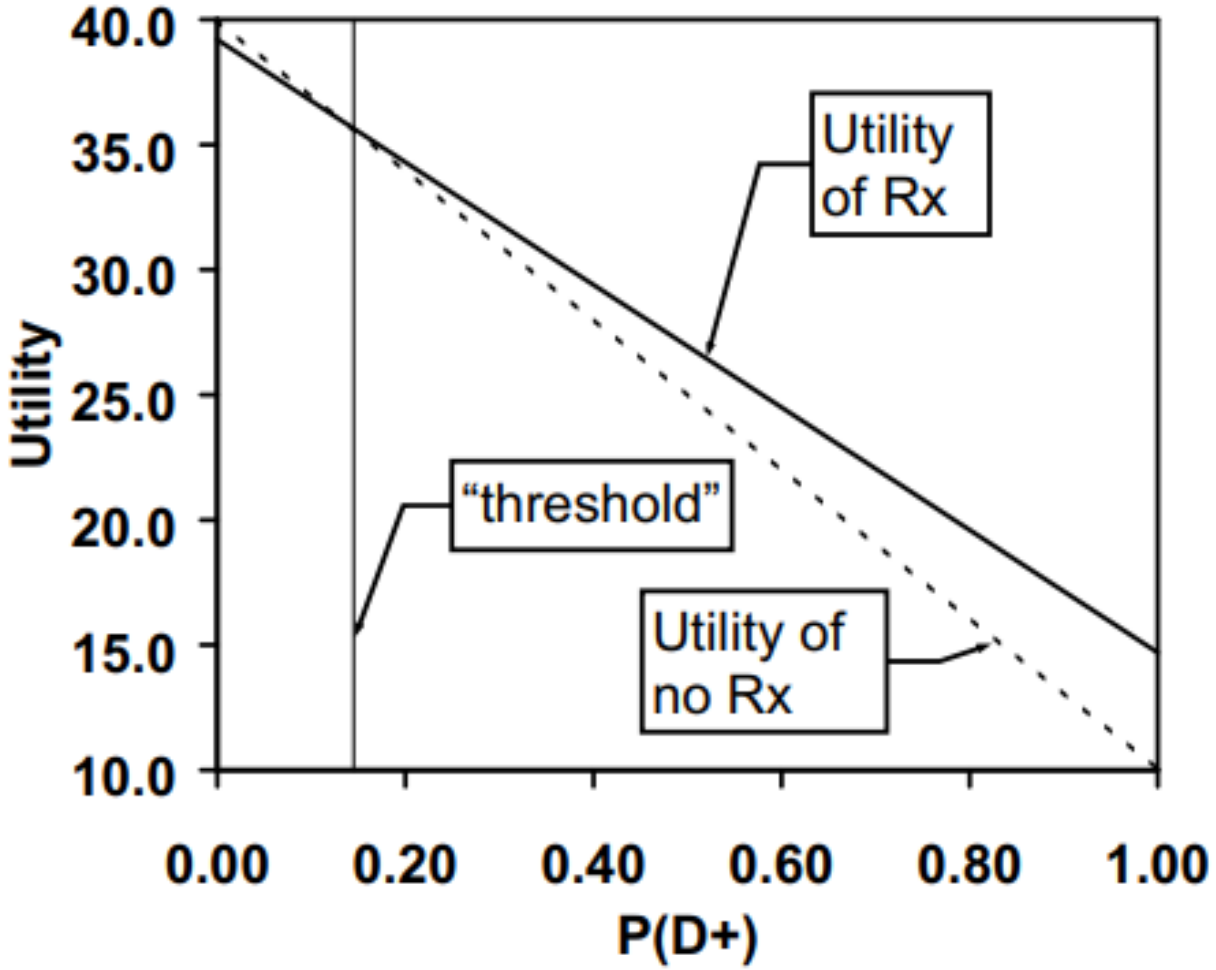
Expected Utility of no treatment:
EU(Rx-) = p(D+) x 10 + p(D-) x 40

Harm of Rx = 40 ***minus*** 39.2 = 0.8 year

Benefit of Rx = 14.7 ***minus*** 10 = 4.7 years

DO WE RECOMMEND OPERATION?

The threshold as a graph



The "threshold" is the value of P(D+) above which the utility of treatment exceeds the utility of no treatment.

This threshold depends on the benefits and the harms as defined before.

Benefits, harms and the threshold

P(D+) = the probability that disease is indeed present
Expected benefit of Rx = ***P(D+)*** x ***benefit***
Expected harm of Rx = ***P(D-)*** x ***harm***

Treat if expected benefit > expected harm, or if
P(D+) x ***benefit > [1 - P(D+)]*** x ***harm,***

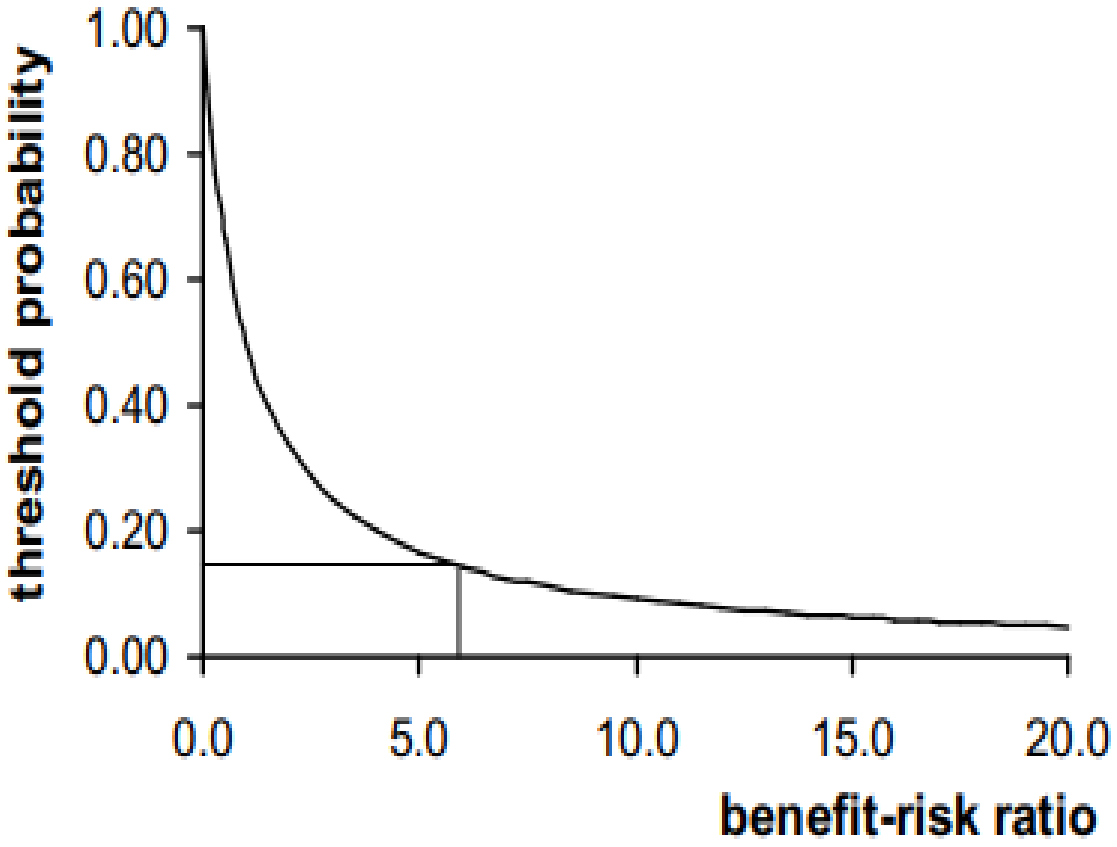
$$P(D+) > \frac{harm}{benefit + harm}$$

This is called the ***threshold probability*** in a treat versus no-treat decision and can also be written as:

$$P(D+) > \frac{1}{benefit/harm + 1}$$

The threshold probability is completely determined by the ratio benefit/harm!

Example of threshold calculation



In our example:
harm = 40 - 39.2 = 0.8 year
benefit = 0.98 x 15 - 10 = 4.7 years
benefit-harm ratio = 4.7/0.8 = 5.9 (rounded)
operate if ***P(D+) > 1 / (5.9 + 1) = 0.14***
(This is the vertical line in the previous graph)

NOTE: ONLY LE CONSIDERED!
SYMPTOMS?

Lessons/conclusions

1. A non-zero probability of disease does not necessarily mean that you treat ***even if treatment is known to be beneficial in patients with disease.***
2. The purpose of diagnostic work-up is to assess whether the probability of disease is ***above, or below, the threshold.***
3. Decisions are “evidence-based” when the estimates of benefits, harms, probability of disease, etc. come from ***sound research*** in ***similar patients.***
4. Clinical decision analysis: way of using the available evidence (if any…)

Deciding when to test

• Before ordering a test ask: What will you do if the test is positive? What will you do if the test is negative? If the answers are the same, then *don't do the test*.

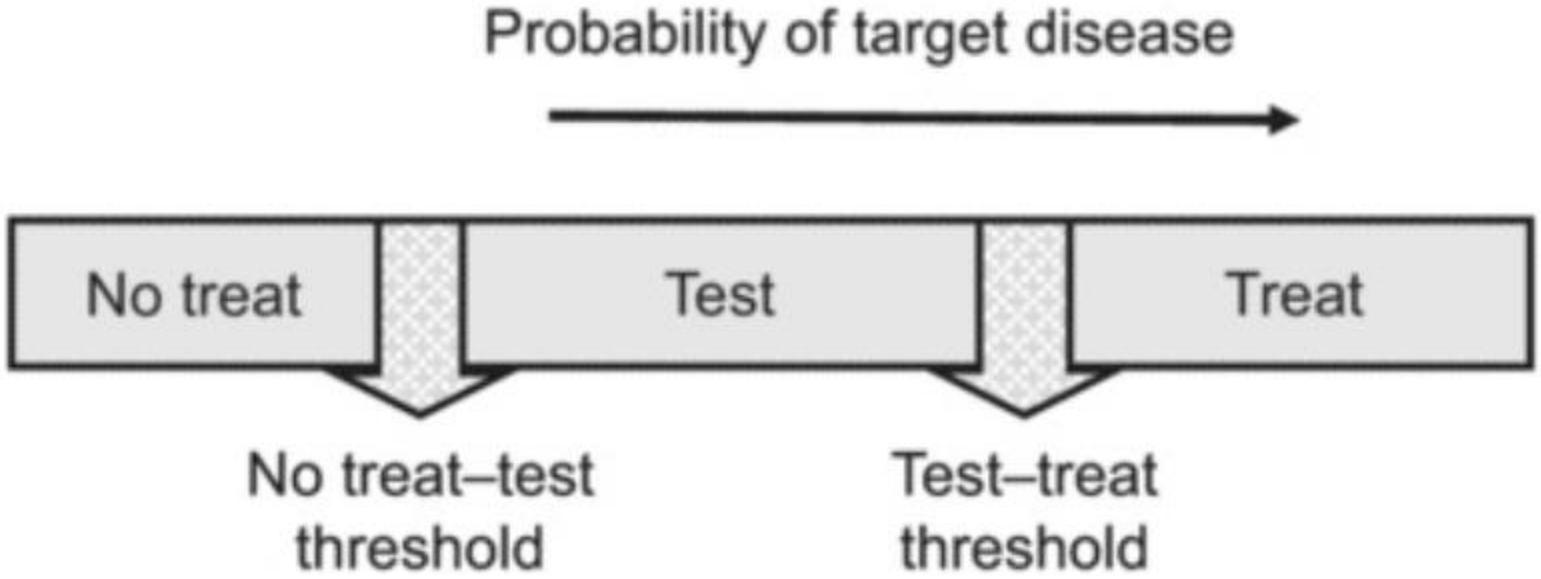
• Whether or not you explicitly calculate the test thresholds, the important concept is that diagnosis is focused around the treatment threshold, and that there is a 'gray zone' around the treatment threshold where testing is worthwhile.

•performing a test to gain additional information is worthwhile only if two conditions hold:

1. at least one decision would change given some test result, and

2. the risk to the patient associated with the test is less than the expected benefit that would be gained from the subsequent change in decision.

Test thresholds: defining the ‘gray zone’

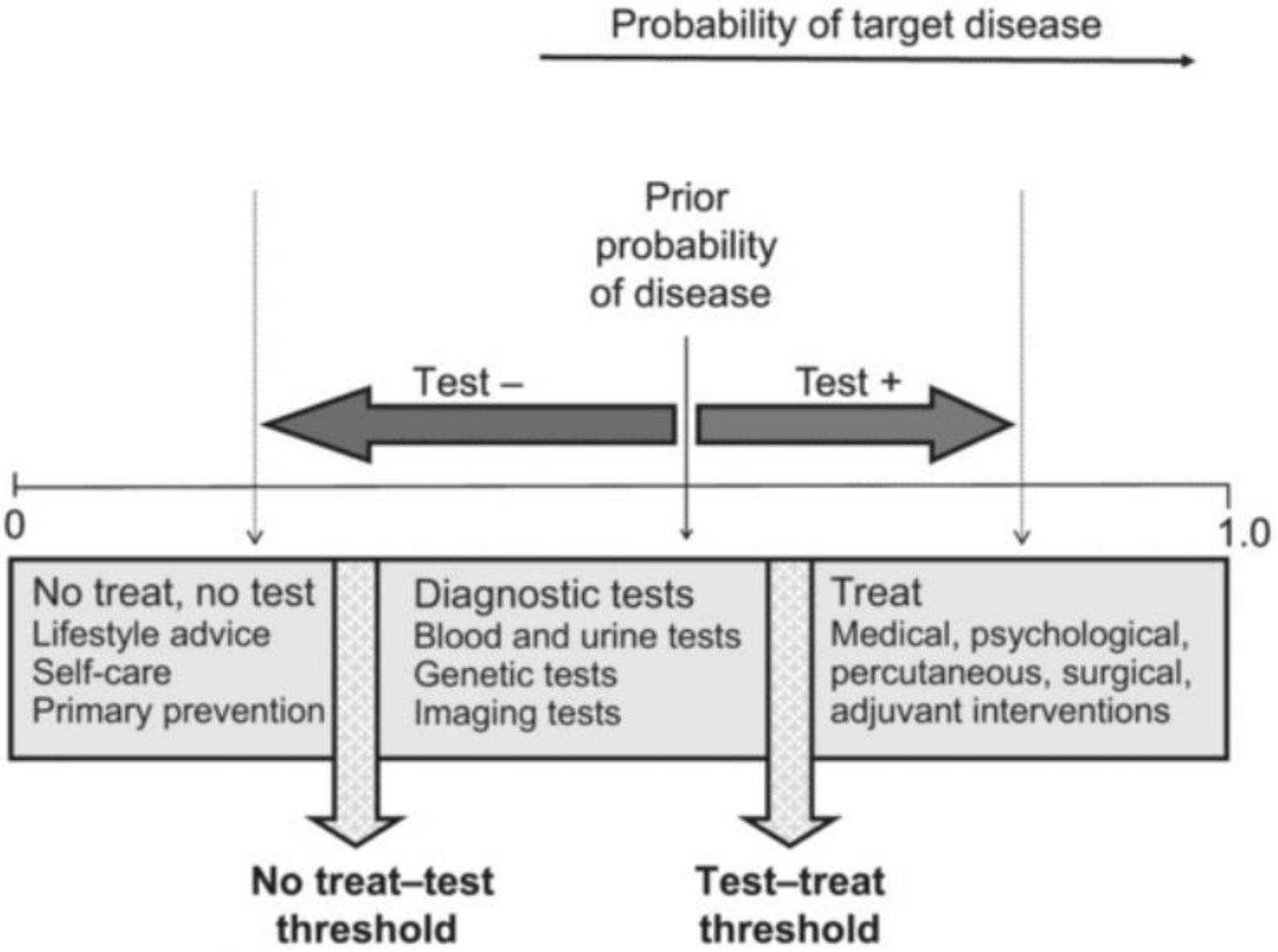


Division of the probability of a disease into three ranges:

(a) do not treat (for the target disease) and do not test, because even a positive test result would not persuade us to treat,

(b) test, because the test will help with treatment decisions; and

(c) treat and do not test, because even a negative test result would not dissuade us from treating



- A test has value if a positive result shifts our decision to the ‘treat’ zone and/or a negative result shifts the decision to the ‘no treat’ zone.

- There will be a range of probabilities around the treatment threshold for which the diagnostic test is capable of changing the choice of treatment. The boundaries defining this ‘gray zone’ of uncertainty are known as the *test thresholds*.

• There are several different ways of calculating the test thresholds:

➢*Method 1*. Use the pre-test post-test graph and the treatment threshold.

➢*Method 2*. Draw the decision tree, then analytically or numerically do a threshold analysis or sensitivity analysis on the probability of disease to find the two thresholds. This is the most general method.

➢*Method 3*. Use the utility graph, which is useful for visualizing the effects of changes in the parameters. You can add a ‘toll’ to allow for test morbidity and mortality conditional on whether the patient has the disease or not.

➢*Method 4*. Use an extended version of the threshold formula. This can give exact values for the thresholds and be readily incorporated in spreadsheet programs.

• Whichever method is used, the central concern is to recognize the existence of the gray zone where testing is useful.

An example: Chest pain – suspected coronary artery disease (CAD)

Imagine you are a primary care physician. A 55-year-old well-educated woman consults you. She recently started having chest pain while running on the beach and during biking against the wind. The pain is substernal and heavy in nature, is only present during very strenuous exercise, and disappears immediately with rest. The chest pain does not bother her in her usual daily activities but she is concerned that she may be at risk for a myocardial infarction (MI) (heart attack) or another cardiovascular disease (CVD) event such as sudden cardiac death, stroke, or a transient ischemic attack (TIA).

She consults you for risk factor assessment and advice on whether to start using medication (statins, beta-blockers, low-dose aspirin) to reduce the risk of having a CVD event. She does not smoke, never has. Her blood pressure is 138/80, heart rate 58, total cholesterol is 4.9 mmol/l (190 mg/dl), HDL is 1.3 mmol/l (50 mg/dl), and BMI is 21 kg/m2. Her family history is negative for MI and stroke but both parents had essential hypertension. She eats a healthy diet and is physically active.

You enter her risk factors in the European SCORE calculator (for fatal CVD events) and the Framingham risk estimator (for fatal and non-fatal heart attacks) and find

that her ten-year risk of having a CVD event (fatal or non-fatal) given her risk factor profile is very low, about 1%. This assumes, however, that her pain is not caused by obstructive coronary artery disease (CAD). If her chest pain is due to CAD the ten-year risk is much higher, about 10%. Optimal medical treatment (OMT) would halve the risk of a CVD event (Relative risk = RR = 50%) irrespective of her baseline risk. Medical treatment, however, carries a ten-year risk of adverse events of about 2% for this patient (serious bleeding due to aspirin use which may be gastrointestinal, epidural/subdural/joint hemorrhage due to (sports) injuries, syncope with serious consequences due to excessive beta-blockade, and myopathy, rhabdomyolysis, or diabetes mellitus due to statins).

•For the purpose of this example we will simplify the problem by assuming that CVD events and adverse events from treatment never occur together and that the patient values CVD events and adverse events from treatment as equally undesirable. Furthermore, we assume the patient wants to maximize her ten-year event-free survival.

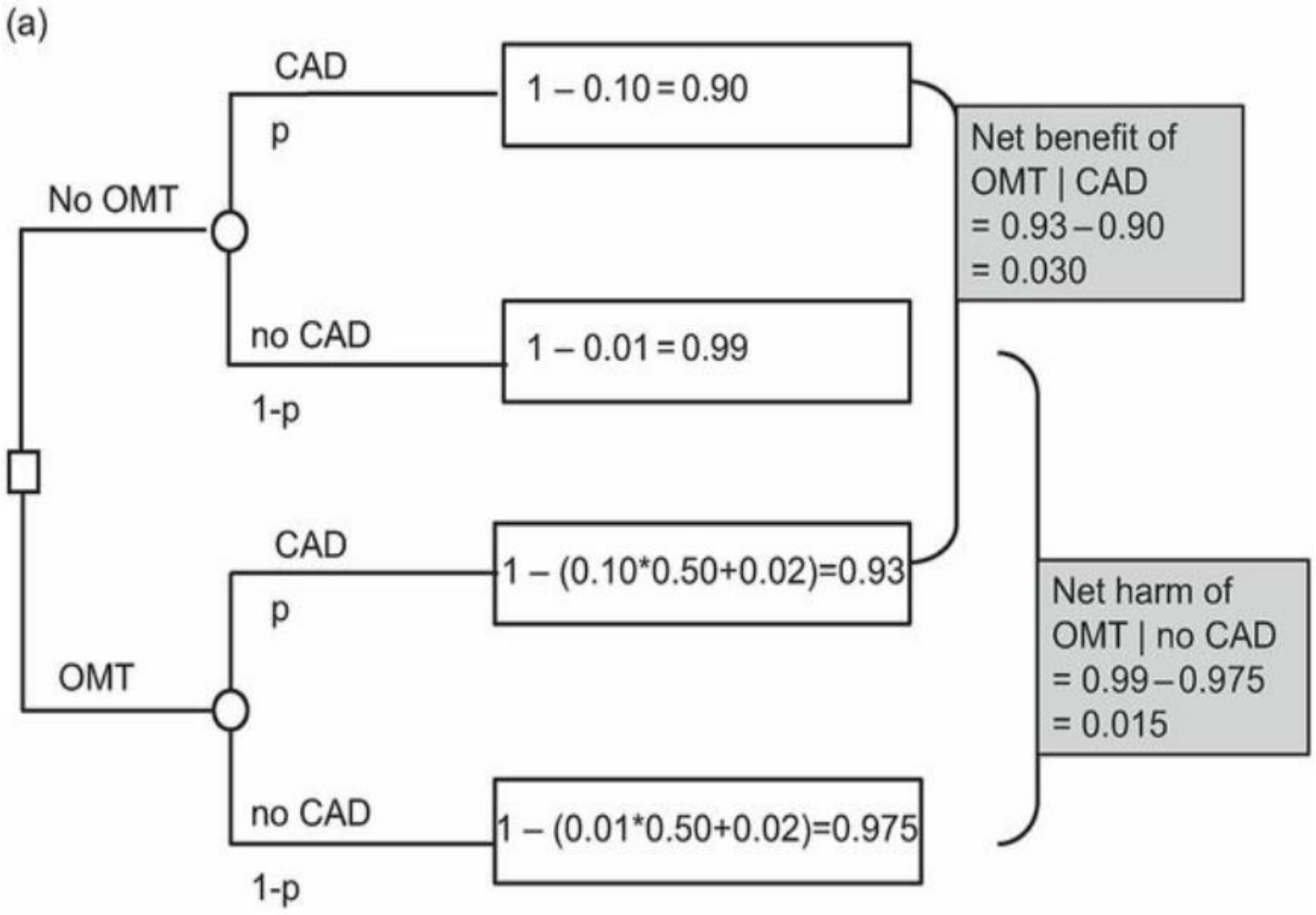
- Clinical balance sheet for alternative management strategies for a patient with suspected coronary artery disease (CAD)

Underlying diagnostic truth	Ten-year outcome	Do not treat	Treat (RR = 0.5)
No CAD	CVD events	0.01	0.5 × 0.01 = 0.005
	Adverse events OMT	0	0.02
	Total events	0.01	0.025
	Event-free survival	**0.990**	**0.975**
CAD	CVD events	0.10	0.5 × 0.10 = 0.05
	Adverse events OMT	0	0.02
	Total events	0.10	0.07
	Event-free survival	**0.900**	**0.930**

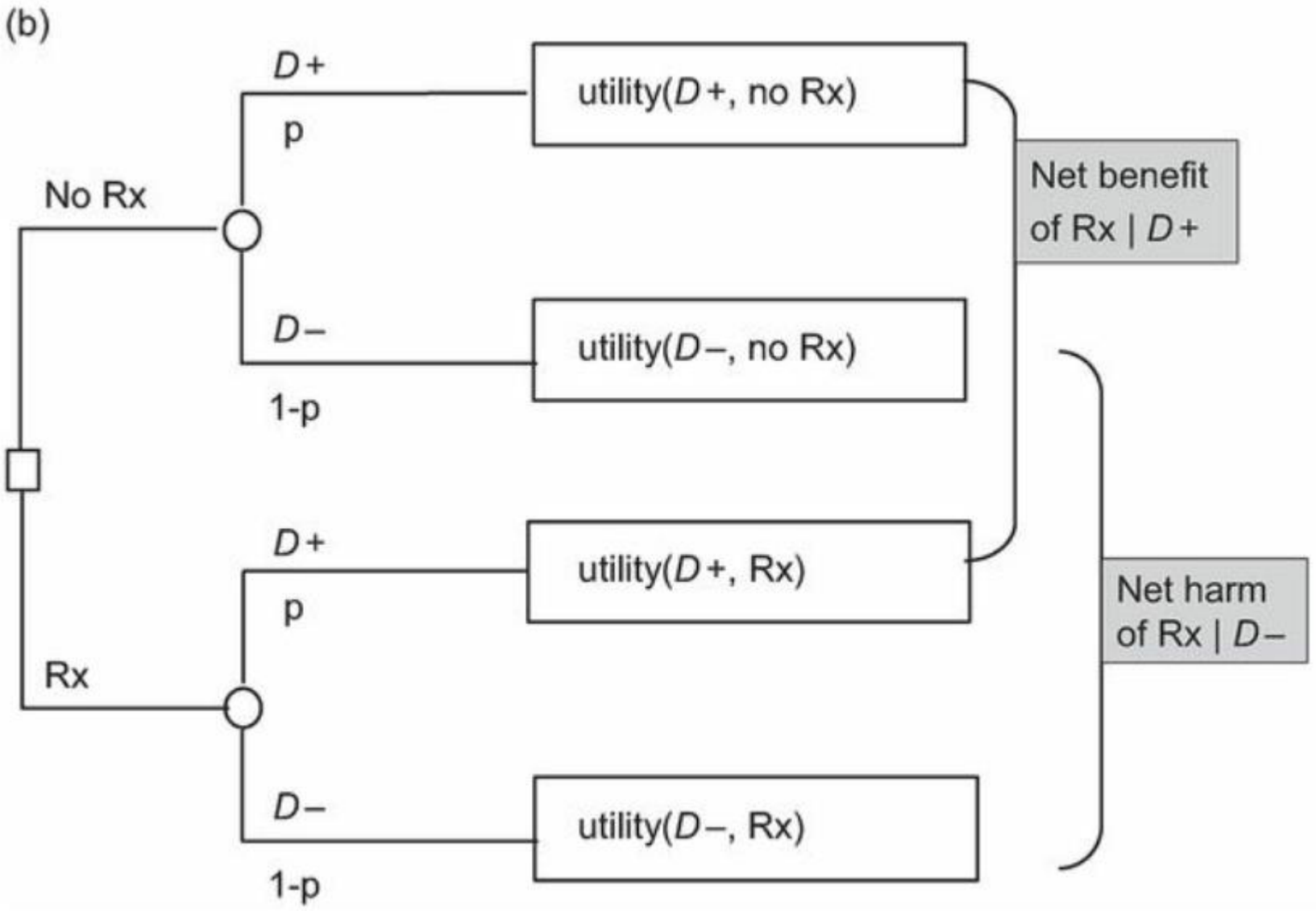
- RR: relative risk
  CAD: coronary artery disease
  CVD: cardiovascular disease
  OMT: optimal medical treatment

•neither strategy is dominant: both have advantages and disadvantages and the decision depends on the prior probability *p* that the patient has the underlying disease (CAD). It will help to calculate the expected value of each option. One way to do this is to draw the decision tree, which will help us structure the sequence of events and probabilities over time.

Decision tree comparing two strategies for (a) suspected coronary artery disease (CAD): optimal medical treatment (OMT) vs. no treatment (no OMT) using ten-year event-free survival as outcome measure
Abbreviations: p prior probability of disease, *D*+ disease present, *D*- disease absent, u utility/outcome measure.



(b) the generic form of the decision tree for a disease (D) comparing treatment (Rx) with no treatment (no Rx).

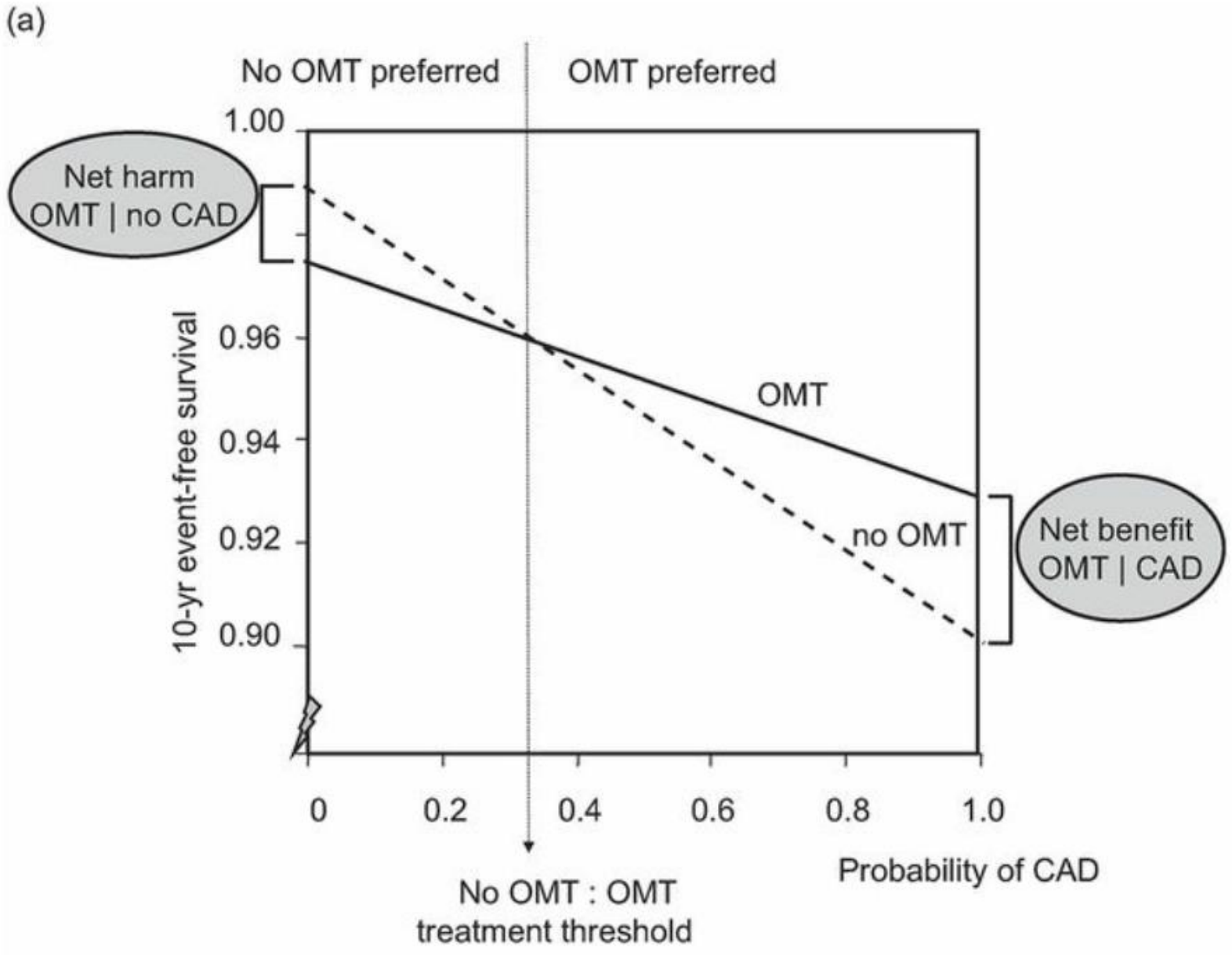


In both (a) and (b) the benefit of treatment conditional on presence of underlying disease and harm of treatment conditional on absence of underlying disease are indicated.

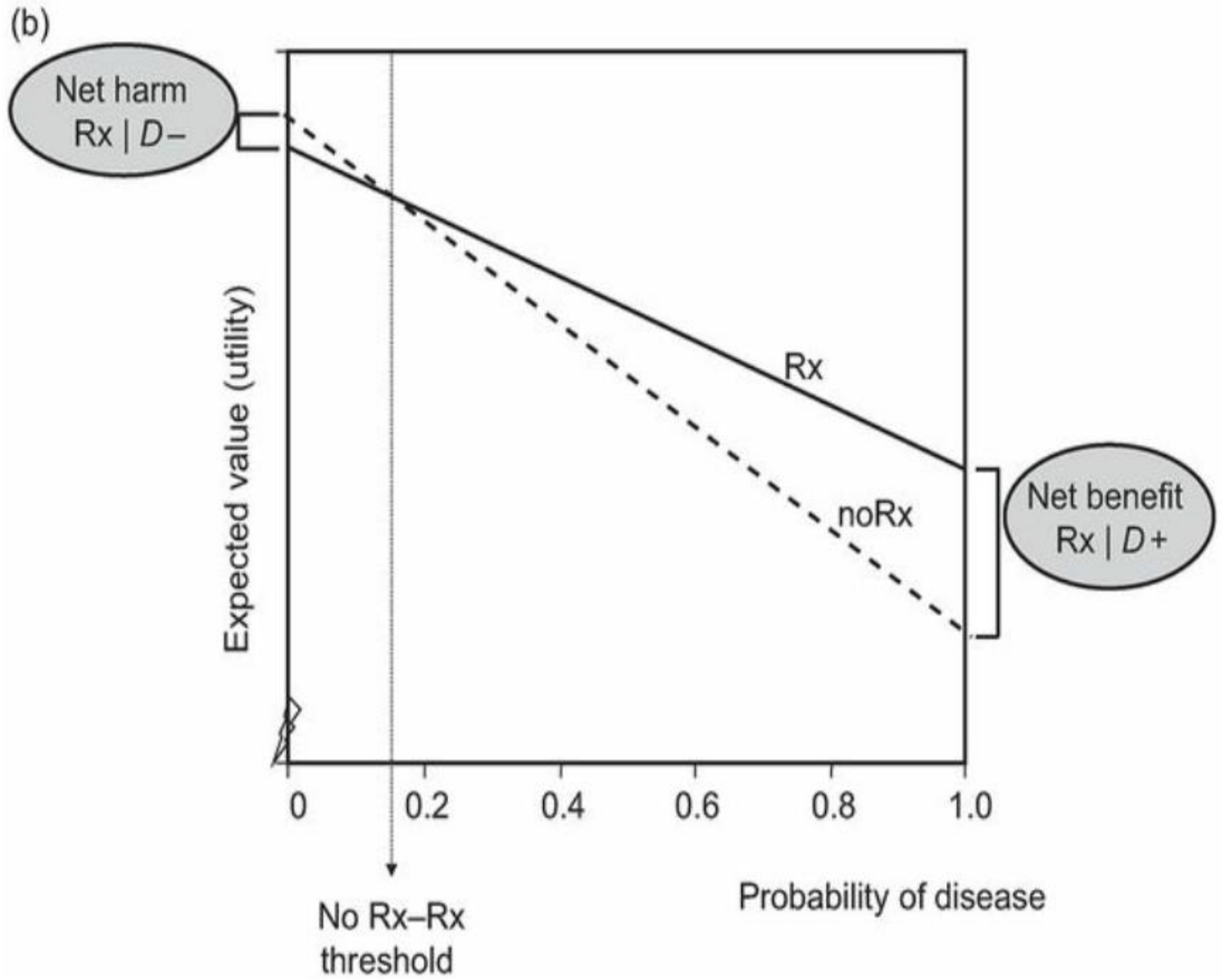
•To determine the treatment threshold we must compare the benefits and harms of treatment vs. no treatment. The benefit of OMT if the patient has CAD is that it reduces the ten-year probability of having a CVD event by 5% (absolute ten-year risk reduction if the patient has CAD). There is, however, a 2% probability of adverse events from treatment. Thus, the net benefit is an absolute ten-year risk reduction of 3% (= absolute increase in ten-year event-free survival of 3%) with treatment compared with no treatment if there is underlying CAD.

•What about the harms to those without CAD? The OMT will also halve their risk of an event, reducing it from 1% to 0.5% but that too would come with a 2% risk of an adverse event from treatment. Thus, the net harm of treating patients without CAD when we should not have treated them (‘leave the well alone’) is 1.5% (= absolute decrease in ten-year event-free survival of 1.5%).

(a) The ten-year event-free survival of preventive treatment (OMT) compared to no treatment (no OMT) as a function of the probability of coronary artery disease (CAD).



(b) The corresponding generic graph for expected utility of treatment (Rx) and no treatment (no Rx) as a function of the probability of disease (D). The net benefit and net harm of treatment compared to no treatment for patients with (*D*+) and without (*D*-) the disease, respectively, have been indicated. No treatment has the highest expected value for low probabilities of disease, whereas treatment has the highest expected value for high probabilities of disease.



•The expected value of treatment and no treatment are equal at the treatment threshold. Note how the treatment threshold shifts depending on the harm to benefit ratio. If the harm to benefit ratio decreases (compare last Figure (b) to (a)), the treatment threshold is lower, broadening the indication for treatment.

$$Treatment\ threshold = \frac{Harm}{Harm + Benefit}$$

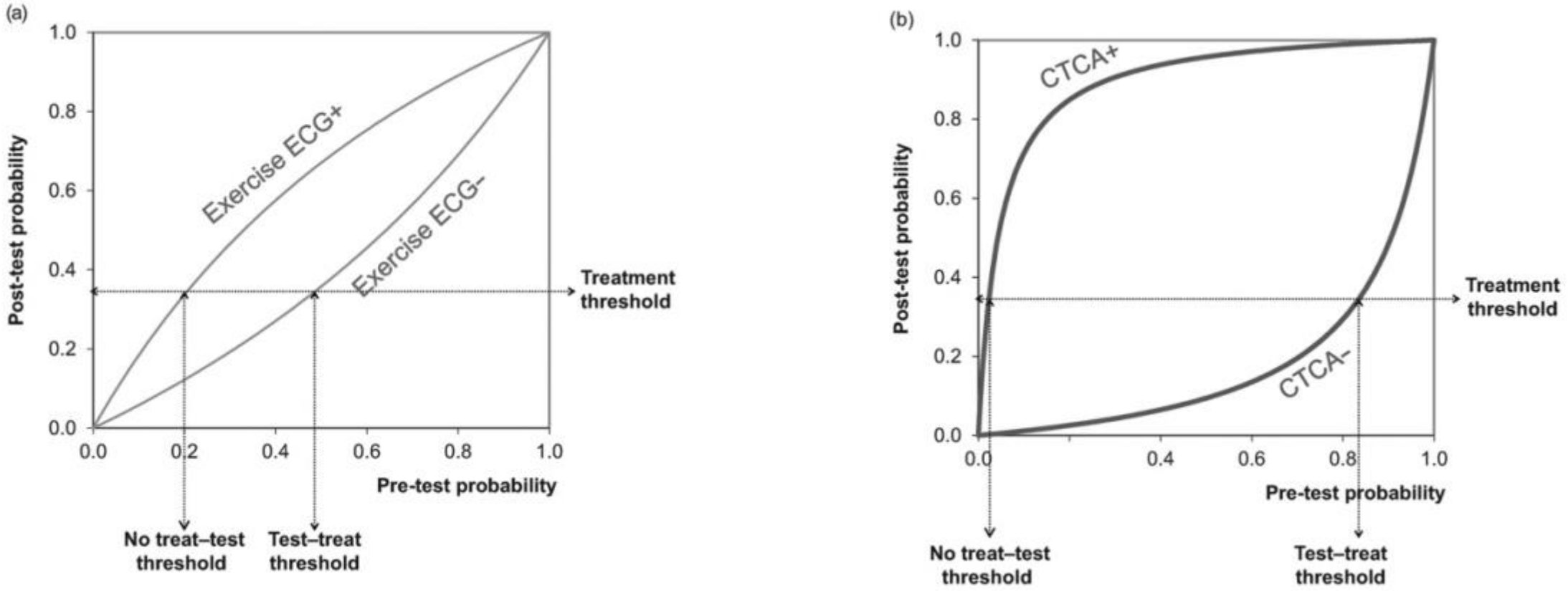
$$Threshold\ p(CAD) = 0.015/(0.015 + 0.030) = 0.33$$

Hence, assuming that we can do no further tests, we would prefer not to treat if the chance of CAD were less than 33% and we would prefer treatment if it were greater than 33%.We can estimate the probability of CAD by consulting a web-based prediction model: based on age, gender, and the type of chest pain (‘typical’ in the example) the probability of CAD is 19%; taking into account that she has no risk factors the probability is 11%. Thus, the probability of CAD is well below the treatment threshold and we would be able to reassure our patient that she should not start medication. A prudent physician would add ‘if you experience persistent or worsening chest pain interfering with your daily activities and reducing your quality of life, you should return for non-invasive diagnostic testing. And keep up the healthy lifestyle!

Example (*cont.*) Testing for suspected coronary artery disease

•After a few months the patient returns with persistent symptoms. A friend her age was recently admitted emergently to hospital with a heart attack. She is worried about her own chest pain and would like to undergo a diagnostic workup.

•Will testing be useful? And which test would we recommend? Traditionally, the most commonly used test for suspected CAD is exercise ECG which, in women, has a sensitivity of 61% and specificity of 70%. There is an alternative test: over the last three decades CT coronary angiography (CTCA) has been developed and has a sensitivity of 90% and specificity of 96%. Let us first calculate the post-test probabilities for all possible pre-test probabilities.



•Graph of the relationship between pre-test and post-test probabilities for the positive and negative test result of (a) exercise electrocardiography (ECG) and (b) CT coronary angiography (CTCA) for suspected coronary artery disease. Superimposing the treatment threshold on the post-test probability (vertical axis) yields the consequent zone of the pre-test probability (horizontal axis) where (a) exercise ECG and (b) CTCA can change the decision. Using a test with higher sensitivity and/or specificity (CTCA rather than ECG) widens the range over which testing is useful.

•When should the results change the treatment decision? The treatment threshold (33%) has been marked on the post-test probability axis, and the zones where testing can change the decision have been indicated. The ability of the test result to change the treatment choice depends not only on its sensitivity and specificity but also on the pre-test probability.

•For example, at a pre-test probability of 30%, a negative test result (both for exercise ECG and for CTCA), would lower the probability of disease, indicating that the decision not to treat is best, whereas a positive test result would increase the probability (to over 33%), suggesting treatment is best.

•The *test–treat threshold* is the probability at which we are indifferent between testing and immediate treatment. It is the probability for which the expected utility of testing and treating is equal.

•At pre-test probabilities above 47%, however, both a negative and positive exercise ECG result would leave us above the treatment threshold, implying that we would choose to treat no matter what the test result, and hence the test does not contribute to the decision. This occurs where the curve of negative test results crosses the treatment threshold. This is the *test–treat threshold*.

•Within the shaded ‘gray zone,’ the test is capable of changing the treatment decision, whereas outside this zone it does not. That is, if you test outside the gray zone and base your treatment decision on the test result, then you would do worse than not testing at all. This is an important point: *Imperfect tests performed inappropriately may do more harm than good because of the subsequent inappropriate treatment decisions*.

•At low probabilities the risk is believing false-positive results, whereas at high probabilities the risk is believing false-negative results. In fact, if the pre-test probability is outside the ‘gray zone,’ and the test has already been performed, a wise decision maker would be better off ignoring the result than be lulled into acting on it (medical–legal considerations notwithstanding)!

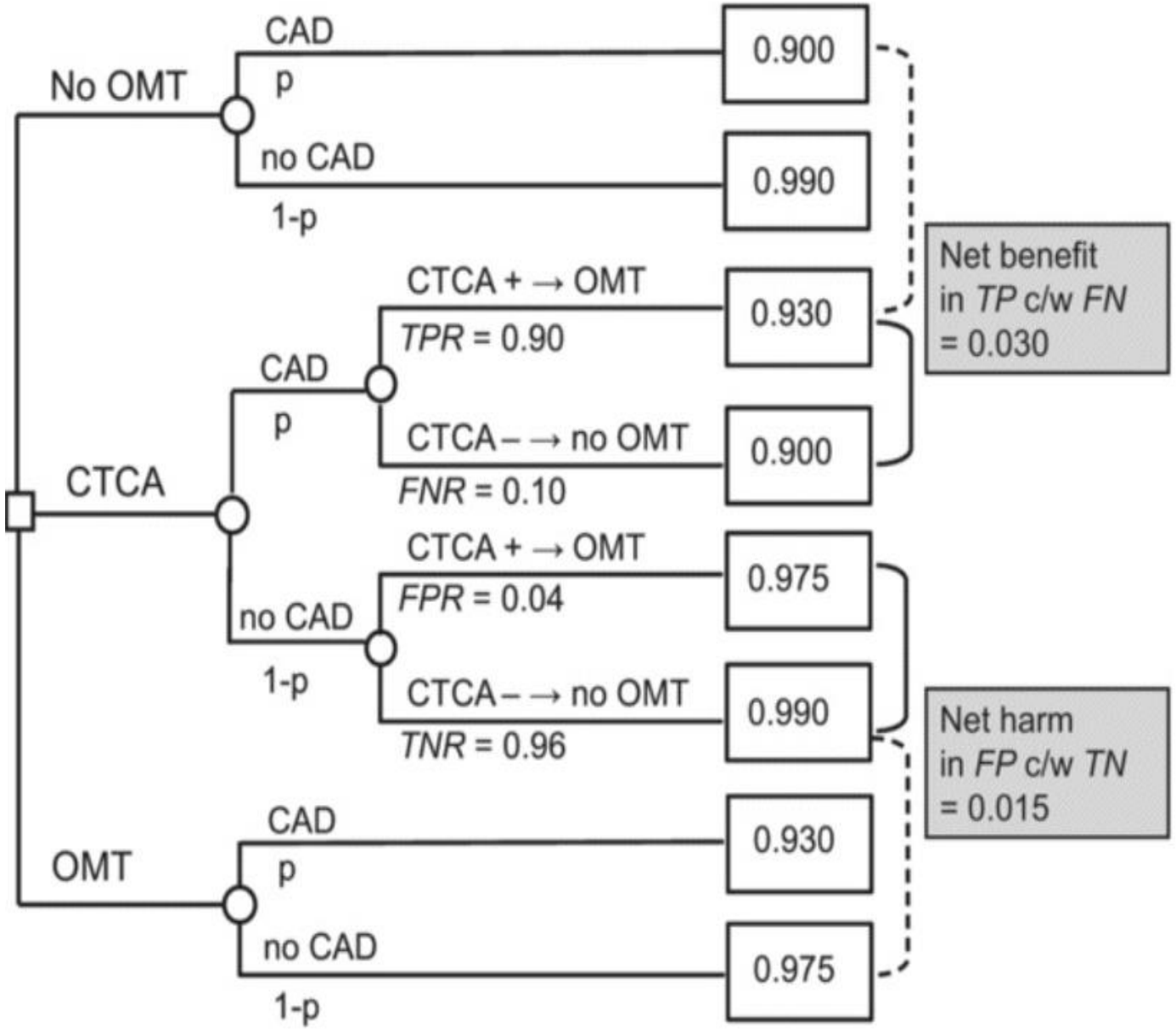
In the example, the gray zone for exercise ECG is 20–47% whereas for CTCA it is 2–83%. Note how the range for testing widens with higher sensitivity and specificity of the test.

•A general solution to finding any threshold is to draw the decision tree and then perform an appropriate threshold or sensitivity analysis. We first modify the benefit vs. harm graph by adding an additional line to represent the test. New figure presents the expected value of the do not treat, test, and treat options for the range of probabilities of disease.

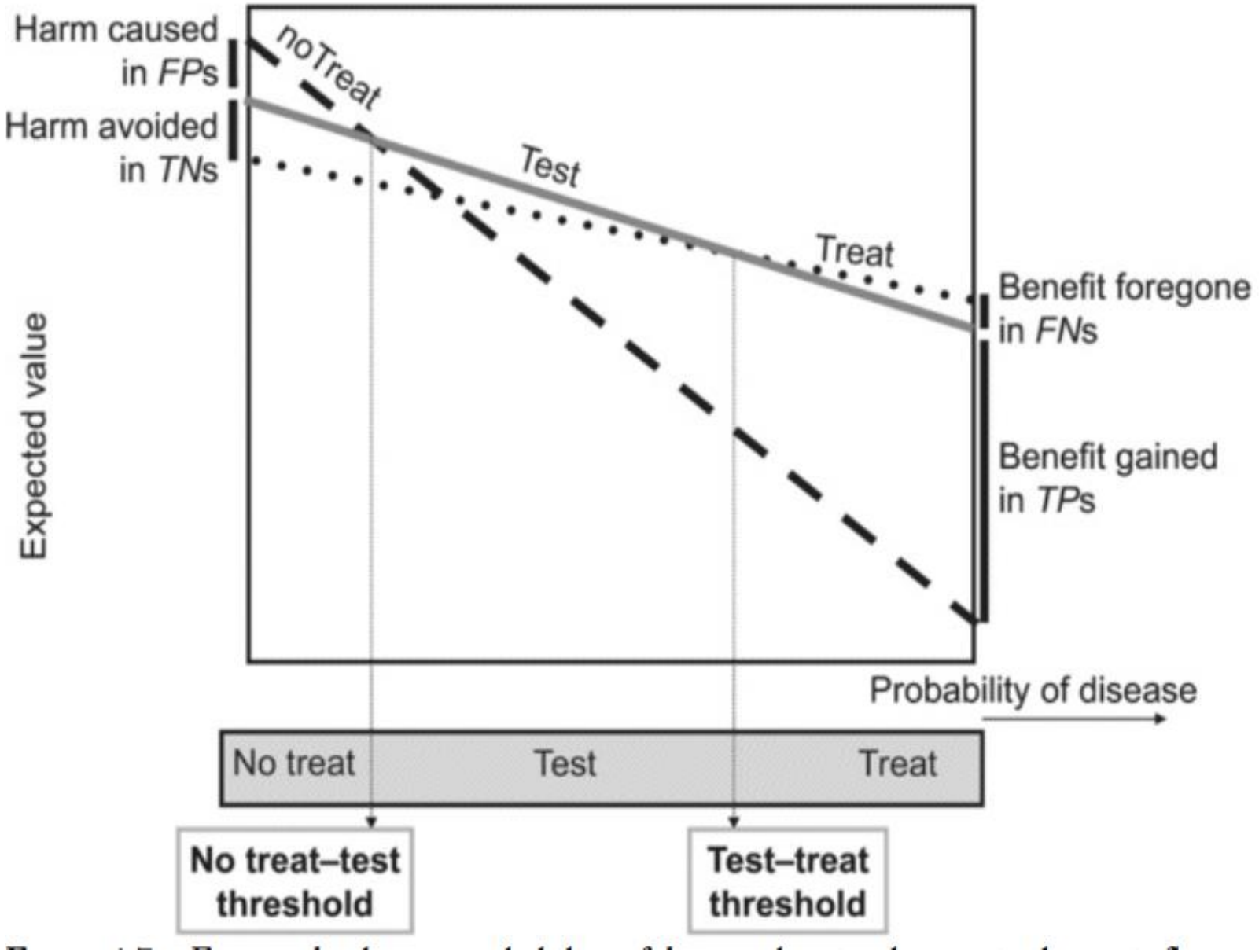
•Notice that the option with the highest utility is optimal: do not treat for low probabilities of disease, treat for high probabilities of disease, and test for intermediate probabilities (the gray zone).

•Note that at this stage we have not yet introduced any ‘toll’ from the test itself.

Decision tree for the choice between no treat–no test (no OMT), test (CTCA), and treat (OMT) options for the case example. Indicated are the net benefit gained in diseased patients correctly identified by the test and treated (*TPR*: true-positive test results) compared with those missed (*FNR*: false negative test results, equivalent to diseased patients not tested–not treated), and also the net harm incurred in non-diseased individuals incorrectly labeled as diseased by the test and treated inappropriately (*FPR*: false positive test results, equivalent to non-diseased individuals not tested but treated) compared with those correctly identified as non-diseased (*TNR*: true-negative test results).



Expected value vs. probability of disease showing how a single test influences the treat vs. do not treat decision. For low probabilities of the disease, do not treat has the highest expected value. For high probabilities of the disease, treat has the highest expected value. In between there is a gray zone in which testing has the highest expected value. (Not drawn to scale, for illustrative purposes.)



- If the graph is drawn to scale, the test thresholds may be read directly from the graph. If the false-positive ratio increases, the ‘test’ line drops particularly on the left-hand side of the graph, increasing the no treat–test threshold.
- Similarly, if the false-negative ratio increases, the ‘test’ line drops particularly on the right-hand side, decreasing the test–treat threshold.
- In both cases the range of prior probabilities across which testing is optimal narrows, implying that there is a more limited indication area for testing.
- Notice that if the false positive ratio is very high and/or the false-negative ratio is very high, testing can even become suboptimal over the entire range of prior probabilities of disease, that is, either ‘no treat’ or ‘treat’ yield a higher expected value, in which case the testing thresholds become meaningless.

Post-test odds = pre-test odds X LR

Odds = prevalence / (1- prevalence)

Prevalence = odds / (1 + odds)

$$No\ treat\text{–}test\ threshold = \frac{Harm \times FPR}{Harm \times FPR + Benefit \times TPR}$$

$$Test\text{–}test\ threshold = \frac{Harm \times TNR}{Harm \times TNR + Benefit \times FNR}$$

Example (*cont.*)

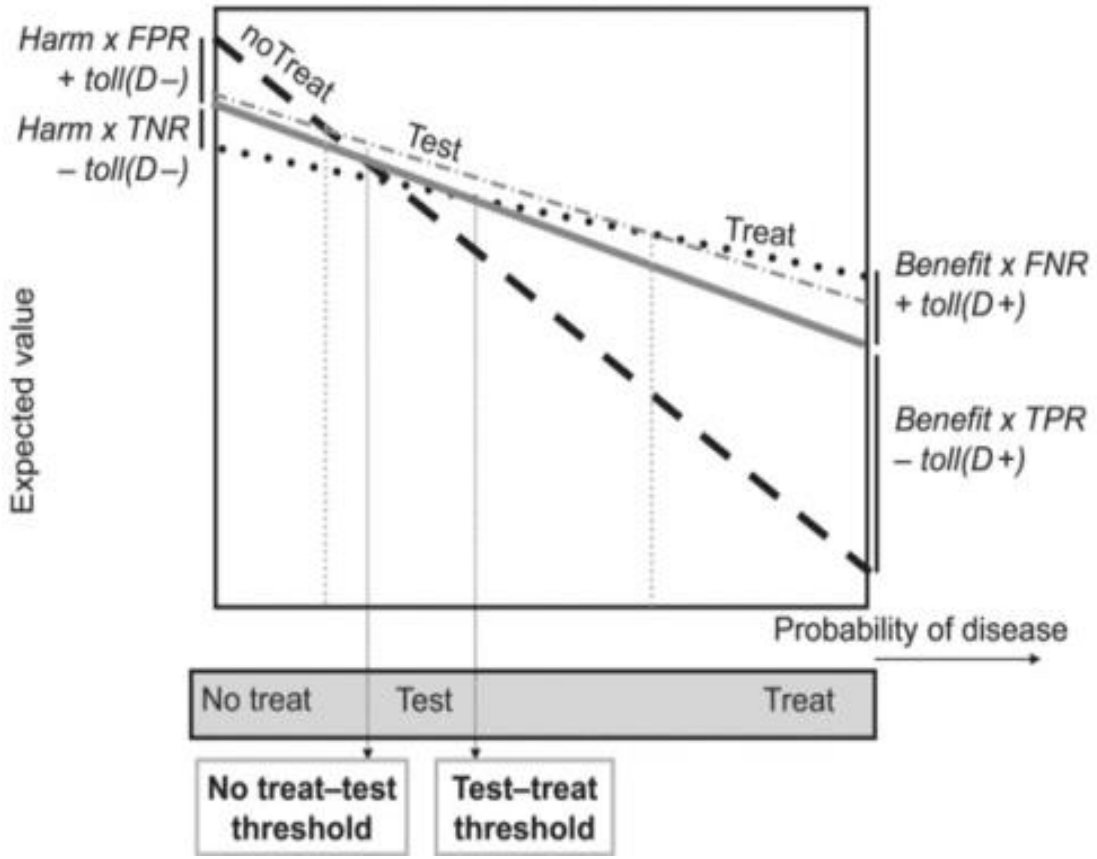
•In the example thus far, we assumed that the exercise ECG and CTCA were available and did not have a risk or cost associated with it. For a first estimate of the usefulness of the test, that is reasonable. In reality exercising a patient who has CAD entails a small risk of inducing a heart attack.

Similarly, CTCA is associated with a risk of a contrast reaction, nephrotoxicity, and radiation-induced cancer.

Thresholds for tests with a ‘toll

$$No\ treat\text{–}test\ threshold = \frac{Harm \times FPR + toll_{D-}}{Harm \times FPR + Benefit \times TPR + toll_{D-} - toll_{D+}}$$

$$Test\text{–}treat\ threshold = \frac{Harm \times TNR - toll_{D-}}{Harm \times TNR + Benefit \times FNR + toll_{D+} - toll_{D-}}$$



•Although the risks of most currently used tests are small, they nevertheless reduce the benefit of testing and need to be considered. Such ‘harms’ of testing narrow the range in which the test is useful; the test thresholds move in towards the treatment threshold.

•The tolls are often unequal, patients with the disease often being at greater risk of adverse events.

The expected value of diagnostic information

•The test thresholds define when a test is useful, but *how useful* is the test in different parts of this range? Particularly near the test thresholds, the incremental gain from testing may be relatively small.

•We need another graphs to quantify precisely the value of the test.