Worked examples of decision curve analysis using Stata

Basic set up

This example assumes that the user has installed the decision curve ado file and has saved the example data sets.

```
use dca_example_dataset1.dta, clear
```

This is a data set giving the results of research study on early detection of cancer. The data include the results of a clinical examination for cancer, the levels of a molecular marker and then the results of biopsy, the "gold standard" for whether or not the patient had cancer.

The results of the clinical exam first have to be turned into a numeric variable.

```
g pos_test = clinical_exam=="suspicious for cancer"
```

We then wish to examine the value of a statistical model that incorporates both the clinical examination and the marker. We create a new variable, giving probabilities of cancer for each patient on the basis of the model.

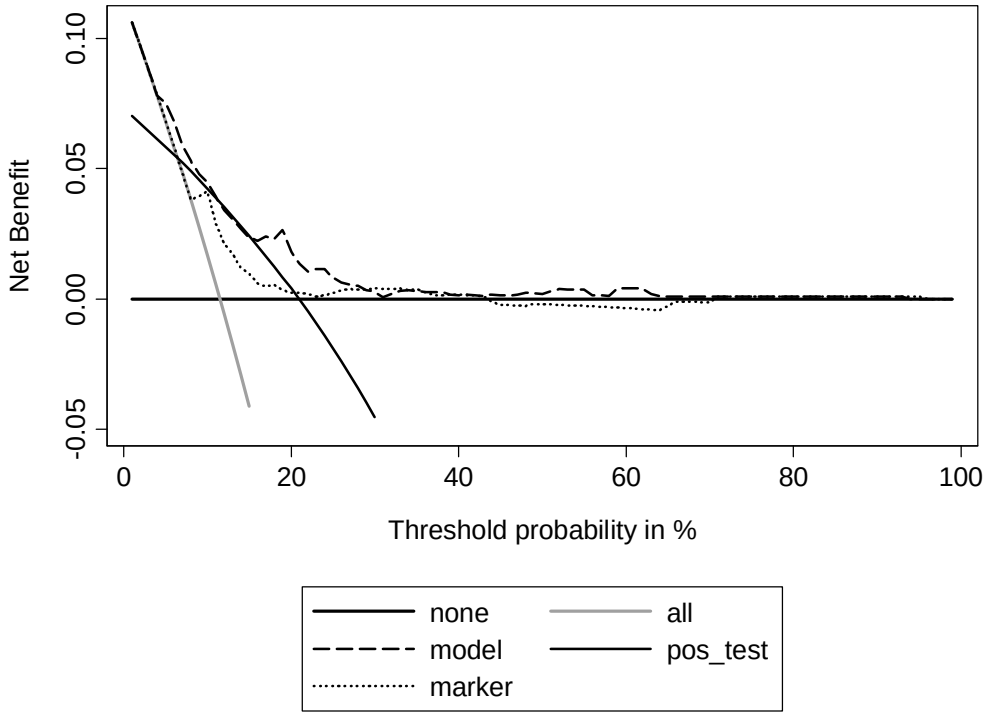
```
logistic cancer marker pos_test
predict model
```

Calculating a decision curve

We now want to compare our different approaches to cancer detection: biopsying everyone, biopsying no-one (i.e. no screening), biopsying on the basis of the clinical exam, biopsying on the basis of the molecular marker, or biopsying on the basis of a multivariable statistical model including both the marker and clinical examination data. We have to specify that the data for the marker variable – unlike those for the model and the test – are not probabilities.

```
dca cancer model pos_test marker, prob(yes yes no)
```

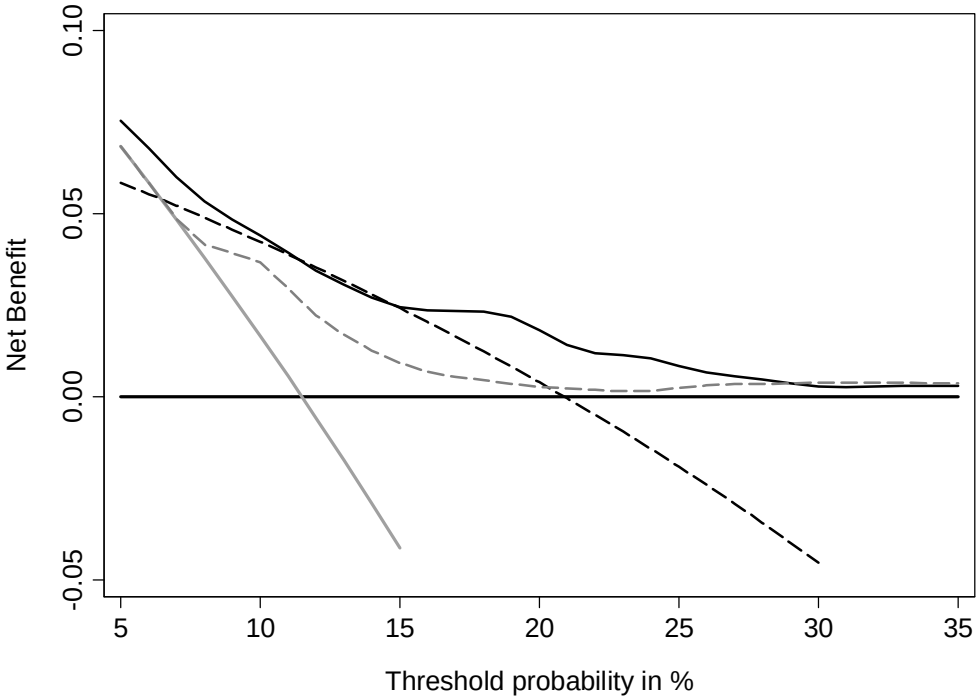
This gives the following graph:



This shows a lot of threshold probabilities that are uninteresting, for example, it is unlikely that someone would require an 80% probability of cancer before they would agree to biopsy. Let's imagine that, after discussing the matter with clinicians, we thought a reasonable range of threshold probabilities was 5% to 35%. In other words, some patients might opt for biopsy even if they had only a 5% chance of cancer, but it would be a rare clinician who would do more than 20 biopsies to find a cancer; on the other hand, few patients would refuse biopsy if their risk was greater than 1 in 3.

Here is a neater way of showing the graph. It focuses on the key range of threshold probabilities (5% - 35%), removes the legend (this can be written in text form), adjusts the line characteristics so that they are easier to read and adds smoothing.

```
dca cancer model pos_test marker, prob(yes yes no) ///
        xstart(.05) xstop(.35) smooth  ///
        clpat(solid dash dash) clcolor(black black gray) legendoff xlabel(5(5)35)
```



The key aspect of the graph is to look at which strategy leads to the highest net benefit. The solid black line corresponds to the model that includes both the marker levels and the results of the clinical exam. It is has the greatest net benefit (i.e. it is the “highest” line) at all threshold probabilities. This is apart from two possible exceptions: between 10 and 15% it is roughly equivalent to using the clinical exam alone; above 30%, it is not much better than using the marker alone. Nonetheless, it is clear that, across the range of reasonable threshold probabilities, one cannot go far wrong basing clinical decisions on the model: it is generally superior, and unlike any alternative strategy, it is never more than trivially worse.

A few general points of interest. First, look at the thin grey line, the net benefit for “treat all”, that is, biopsy everyone. This crosses the y axis at the prevalence. Imagine that a man had a risk threshold of 12%, and asked his risk under the "biopsy all" strategy. He would be told that his risk was the prevalence (12%). When a patient’s risk threshold is the same as his predicted risk, the net benefit of biopsying and not biopsying are the same. Second, the decision curve for the binary test (the clinical exam, black dashed line) crosses that for “biopsy all men” at 1 – negative predictive value and again, this is easily explained: the negative predictive value is 94%, so a patient with a negative test has a probability of disease of 6%; a patient with a threshold probability less than this – for example, a patient who would opt for biopsy even if risk was 3% - should therefore be biopsied even if the clinical exam is negative. The decision curve for the binary test is equivalent to biopsy no-one at the positive predictive value. This is because for a binary test, a patient with a positive test is given a risk at the positive predictive value.

Saving out net benefit

If we wanted to show the net benefits in a table, we could save them out. The following code can be used to produce a table giving the net benefits of each strategy, and the advantage of the model, at risk thresholds of 5%, 10%, 15% … 35%.

```
dca cancer model, xstart(0.05) xstop(0.35) xby(0.05) ///
        saving("decision curve output.dta") replace
use "decision curve output.dta", clear
g delta = modelp1 – all
label var delta “Increase in net benefit from using statistical model”
```

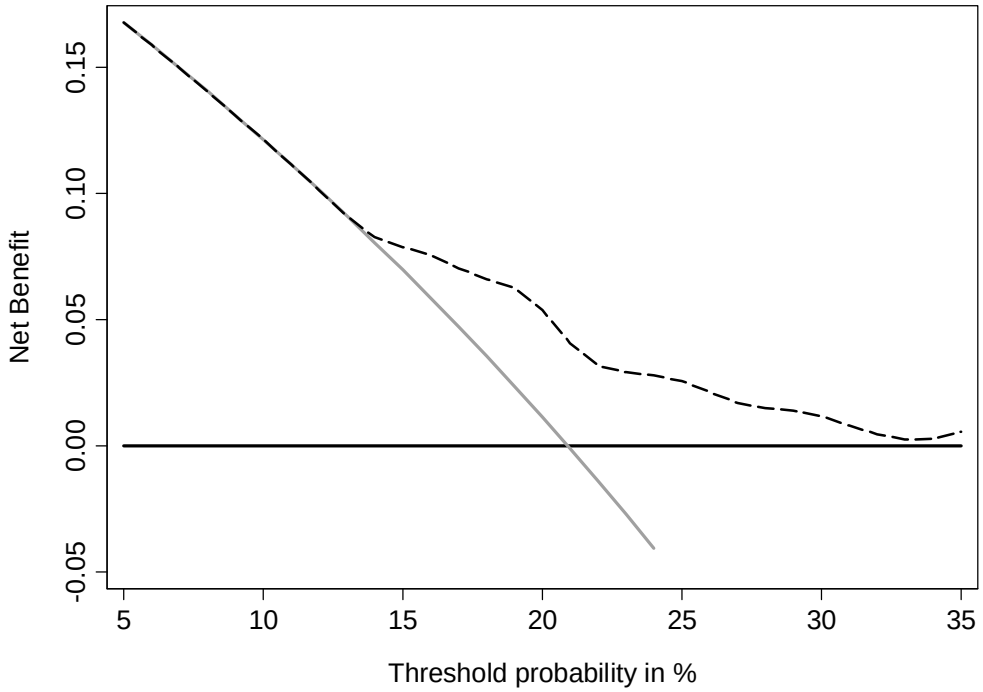
Showing net reduction in interventions

Now let’s imagine that the traditional approach was to biopsy everyone suspicious for cancer on clinical examination. We want to know whether assessing the marker in these patients would help reduce unnecessary biopsies. A typical research study to address this question would only have biopsy data on patients with a suspicious clinical examination. We’ll first create this data set.

```
use dca_example_dataset1.dta, clear
keep if clinical_exam=="suspicious for cancer"
```

Here is the decision curve.

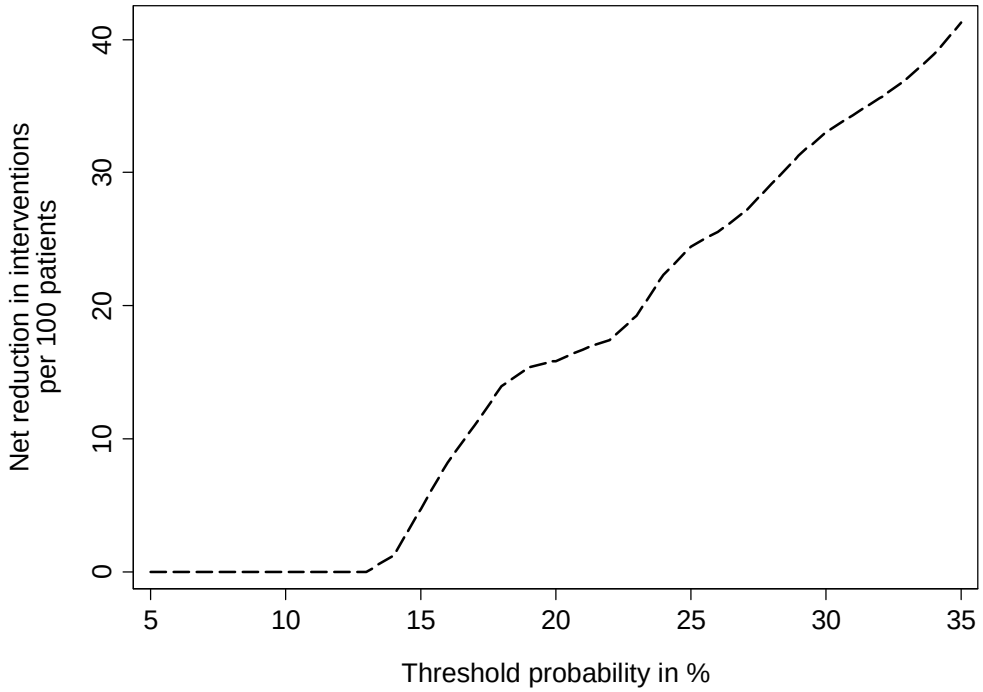
```
dca cancer marker, prob(no) xstart(.05) xstop(.35) smooth legendoff xlabel(5(5)35)
```



Use of the marker has superior net benefit from a threshold probability of about 15% or more. This suggests that measuring the marker would improve outcome for all but the most risk averse patients. The appropriate clinical conclusion would depend on the expense of measuring the marker, and the estimated prevalence of very risk averse patients. For example, if the marker were inexpensive, and few patients were willing to have biopsy at very low probabilities of cancer, it would be reasonable to measure the marker on all patients. If, on the other hand, the marker was expensive, or many patients were thought to have low threshold probabilities, a reasonable approach would be to conduct the clinical exam, discuss the results with the patient, and then order the marker if the patient was unsure what to do.

Net benefit has a ready clinical interpretation. In the figure above, the net benefit of 0.05 at a threshold probability of 20% can be interpreted as: "Comparing to conducting no biopsies, biopsying on the basis of the marker is the equivalent of a strategy that found 5 cancers per hundred patients without conducting any unnecessary biopsies." However, the routine clinical practice is to biopsy everyone with a suspicious clinical examination and what we want to know is the effect on the number of unnecessary biopsies. We can show this using decision curve analysis by specifying **inter**vention.

```
dca cancer marker, prob(no) xstart(.05) xstop(.35) ///
    smooth legendoff inter xlabel(5(5)35)
```



At a probability threshold of 20%, the net reduction in interventions is about 18 per 100 patients. In other words, at this probability threshold, biopsying patients on the basis of the marker is the equivalent of a strategy that reduced the biopsy rate by 18%, without missing any cancers.

Joint or conditional tests

Many decisions in medicine are based on joint or conditional test results. A classic example is where patients are categorized on the basis of a test as being at high, low or intermediate risk. Patients at high risk are referred immediately for treatment; patients at low risk are reassured and advised that no further action is necessary; patients at intermediate risk are sent for an additional test, with treatment decisions made accordingly. Alternatively, the results of two tests can be combined using an algorithm, rather than a multivariable model.

Decision curve analysis can incorporate joint or conditional testing. All that is required is that appropriate variables are calculated from the data set; decision curves are then calculated as normal. To see how this works in practice, see the following example.

```
use dca_example_dataset2.dta, clear
```

These data are from a heart disease study, with data on whether the patient had a subsequent cardiovascular event (yes or no), the result of a clinical exam (high low or intermediate risk) and the results of a scan. There are five clinical options:

1. Treat everyone with medication

2. Treat no-one (i.e. no screening)
3. Treat everyone at high risk from the clinical exam; don't do a scan
4. Scan everyone, then treat everyone who is *either* high risk from the clinical exam *or* who has a positive scan (i.e. joint approach)
5. Treat everyone at high risk; scan patients at intermediate risk and treat those with a positive scan (i.e. conditional approach)

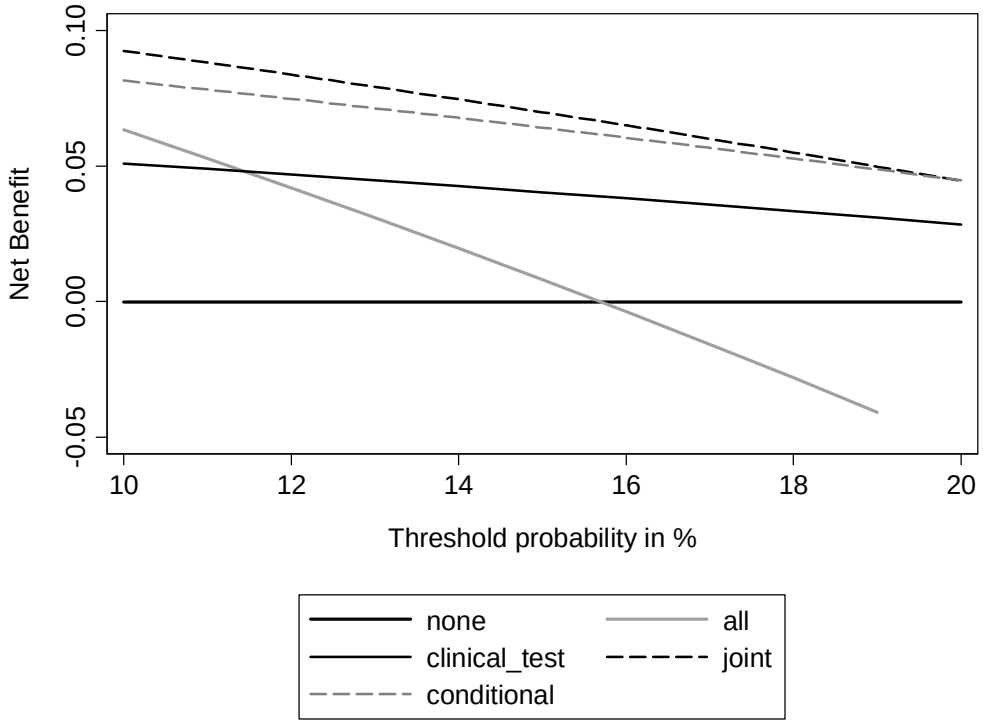
Here is how to create the variables to use in decision curve analysis.

```
*clinical exam: treat high risk patients only
g clinical_test= clinical_exam=="high risk"
*joint test is positive if either:
* scan is positive or
* clinical exam gives high risk
g joint=clinical_exam=="high risk" | scan==1
*conditional test: treat if high risk; scan if intermediate risk
g conditional=clinical_exam=="high risk" ///
        | clinical_exam=="intermediate risk" & scan==1
```

Now we have the variables worked out, we can run the decision curve analysis. In cardiovascular disease, treatment thresholds are typically 10% - 20.

```
dca event clinical_test joint conditional,  ///
        xstart(.1) xstop(.2) ///
        clpat(solid dash dash solid) clcolor(black black gray gray)
```

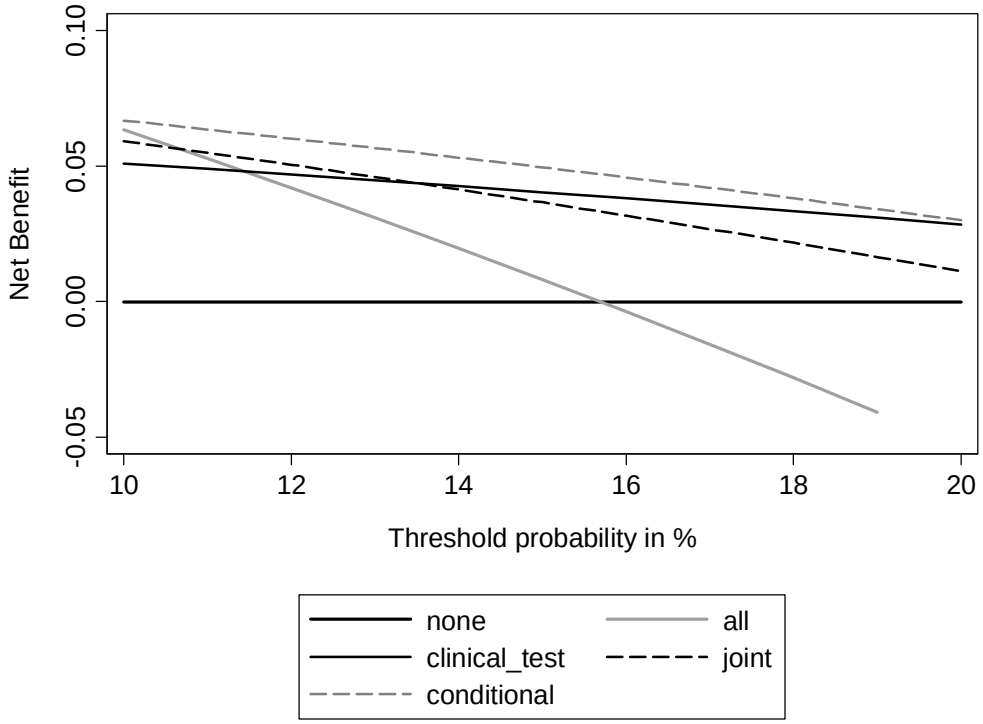
This gives the figure below:



This appears to show that the joint test is the best option, because it has the highest net benefit across almost the whole range of threshold probabilities. The obvious disadvantage of the joint test is that everyone needs to be scanned, and scanning is expensive and time consuming. To incorporate the harm of scanning, we ask a clinician, who tells us that, even if the scan was perfectly accurate, few clinicians would conduct more than 30 scans to predict one cardiovascular event. The "harm" of the scan is the reciprocal of 30, or 0.0333.

To construct the decision curves for each strategy we now incorporate harm. We have to calculate harm specially for the conditional test, because only patients at intermediate risk are scanned.

```
*the harm of a scan is stored in a local
local harm_scan =0.0333
*in the conditional test, only pts at intermediate risk are scanned
g intermediate_risk= clinical_exam=="intermediate risk"
sum intermediate_risk
*harm of the conditional strategy is:
*proportion scanned multiplied by harm of the scan
local harm_conditional = r(mean)*`harm_scan'
*now run the decision curve analysis
dca event clinical_test joint conditional, ///
        harm(0 `harm_scan' `harm_conditional') ///
        xstart(.1) xstop(.2)  ylabel(-0.05(0.05)0.1, format(%9.2f)) ///
        clpat(solid dash dash solid) clcolor(black black gray gray)
```

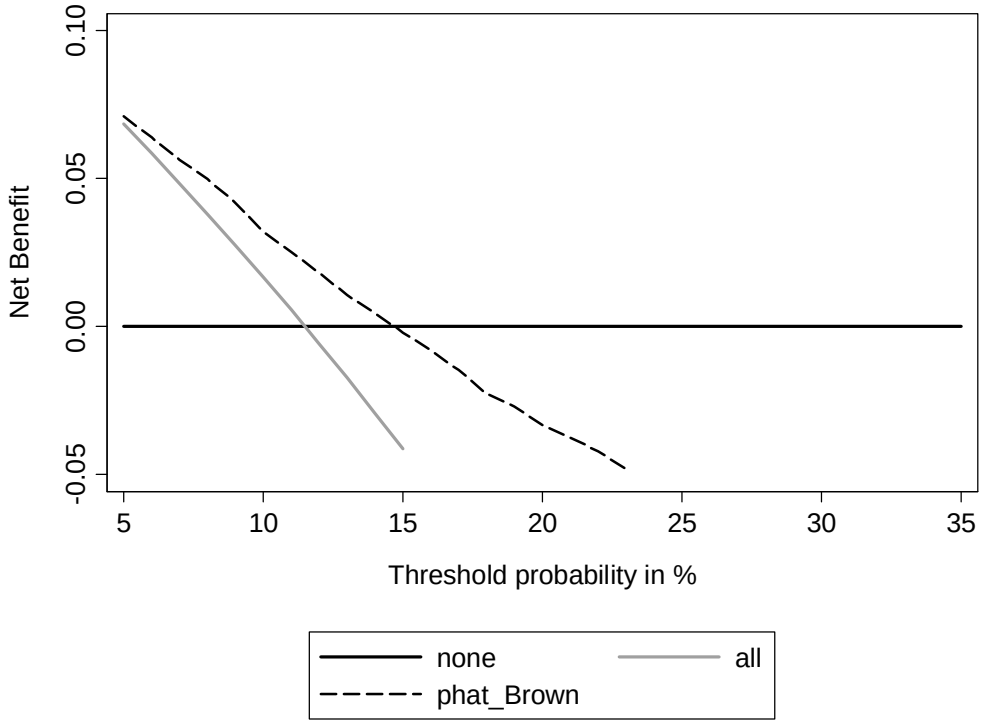


Here the conditional test is clearly the best option, in fact, once you take into account the harm of scanning, it is clearly not worth scanning everyone: the net benefit of just using the clinical exam is often higher than that of the joint test.

Evaluation of published models

Imagine that a model was published by Brown et al. with respect to our cancer biopsy data set. The authors reported a statistical model with coefficients of 1.5 for a positive clinical exam; 0.042 for each 10 point increase in the marker, and an intercept of -3.5. To test this formula on our data set:

```
use dca_example_dataset1.dta, clear
g logodds_Brown=1.5*(clinical_exam=="suspicious for cancer")+ ///
        0.42 * (marker/10) - 3.5
g phat_Brown=invlogit(logodds_Brown)
dca cancer phat_Brown, xstart(.05) xstop(.35) xlabel(5(5)35)
```



This decision curve suggests that although the model might be useful in the most risk averse men, it is actually harmful in patients with more moderate threshold probabilities. As such, the Brown et al. model should not be used in clinical practice.

Application to case-control data

The problem with applying decision curve analysis to case-control data is that net benefit depends on prevalence, and prevalence in case-control studies is fixed by design. This problem can be solved by use of recalibration (sometimes called a "Bayes factor").

```
use dca_example_dataset3.dta, clear
```

This is a case-control study to assess whether a certain genetic mutation, along with smoking history, can predict a cancer. The matching was 3:1, such that the prevalence of cancer in the study was 25%. We will assume that the true risk of cancer by age 75 was 5%. We will first build a model using the data set and then obtain the log odds of disease (linear predictor or fitted values).

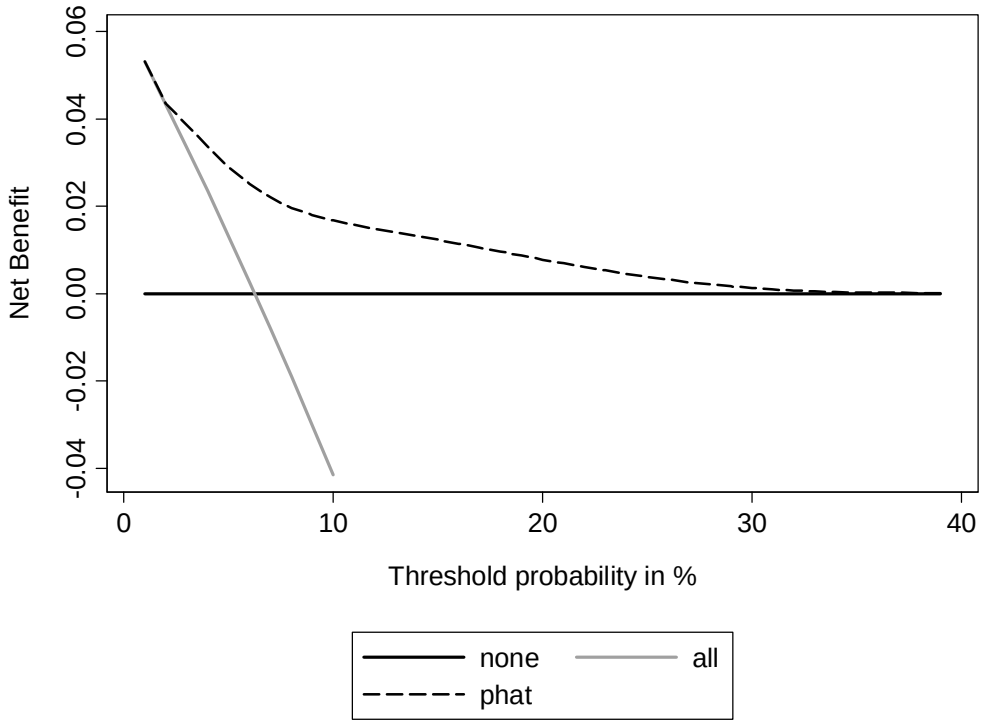
```
logit cancer gene packyears
predict xb, xb
```

We can then add the Bayes factor to the linear predictor, which is the log of the true odds of cancer divided by the odds of cancer in the case-control study.

```
local true=0.05
sum cancer
local design=r(mean)
local Bayes=log((`true'/(1-`true')) / (`design'/(1-`design')))
replace xb=xb +`Bayes'
```

This is then converted to a probability. We use this probability both as the predictor and the outcome for the decision curve, using the assumption that the model is well-calibrated.

```
g phat=invlogit(xb)
dca phat phat, xstop(.4)
```



This decision curve can be used to evaluate whether a statistical model based on smoking history and gene status should be used determine intensity of cancer screening. The decision curve shows that the statistical model would be helpful for decisions with a threshold probability of 2% or above. In other words, in the unlikely event that a patient would agree to be screened even if he or she had a 1% lifetime risk of cancer, they should be screened

irrespective of smoking history and gene status; otherwise, screening should be based on the risks from the model.

Application to survival data

Again, we'll assume that the user has saved the stdca.ado file and the example data sets. We also assume that the user has downloaded stcompet.ado, which is required for the competing risk analysis.

```
use dca_example_dataset4.dta, clear
```

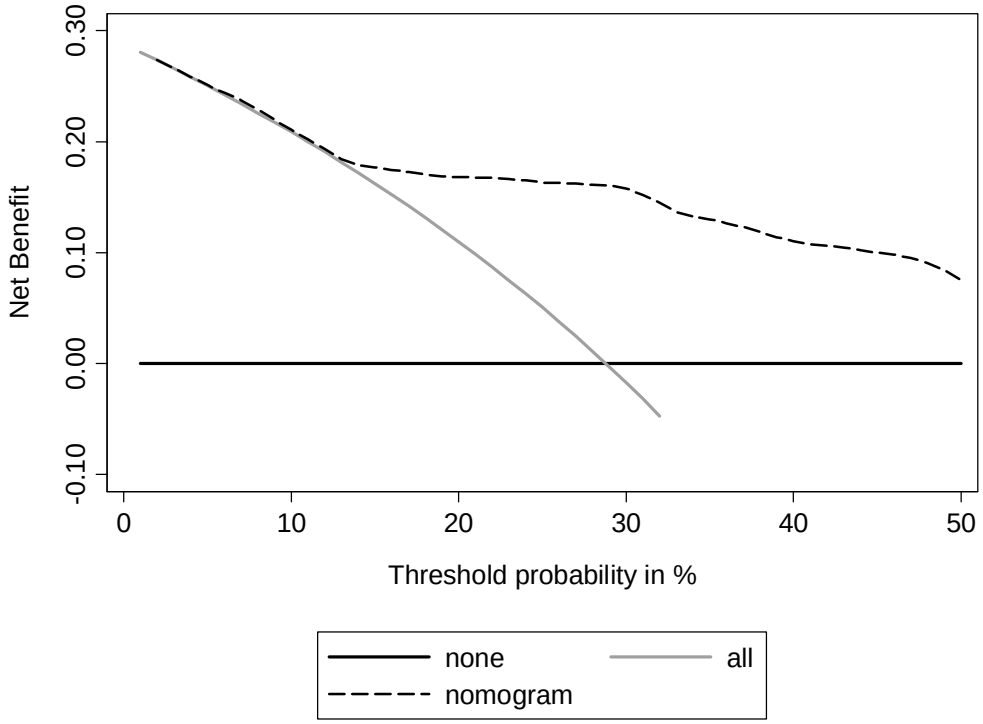
This is a bladder cancer data set consisting of seven variables: some patient characteristics (age at surgery, pathologic stage, and nodal status), a time to event variable in months, whether the patient recurred, whether the patient died, and then predicted probability of recurrence by 60 months, obtained from a statistical model (or "nomogram").

We first need to stset the data, with recurrence considered the failure event.

```
stset ttlastfollowrec, f(recurrence)
```

The code for running the decision curve analysis is straightforward. All we have to do is specify the timepoint we are interested in. Here we ask: how good is the statistical model at predicting recurrence by five years? We'll use a timepoint of 60, which corresponds to 60 months or 5 years in this data set.

```
stdca nomogram, timepoint(60)  smooth xstop(.5)
```



This shows that using the nomogram to inform clinical decisions will lead to superior outcomes for any decision associated with a threshold probability of above 5% or so.

The input predictor variables for stdca must be given as probabilities, even when the prediction model consists of a single variable. In the previous example, the nomogram probability was calculated from a published statistical model. Let's say we wanted to build a model of our own, based on age, pathologic stage, and nodal status. The survival probability to time-point t can be derived from any type of survival model; here we use a Cox as this is the most common model in statistical practice. The formula for a survival probability from a Cox model is given by:

$$s(t|X) = s_0(t)^{\exp(XB)}$$

Where X is matrix of covariates in the Cox model, B is a vector containing the parameter estimates from the Cox model, and $s_0(t)$ is the baseline survival probability to time t.

This can be done in Stata by running a Cox model with age, pathologic stage, and nodal status as predictors, saving out the baseline survival function in a new variable called s0, and obtaining the linear prediction from the model for each subject.

```
stcox age pathstage nodalstatus, basesurv(s0)
predict xb, xb
```

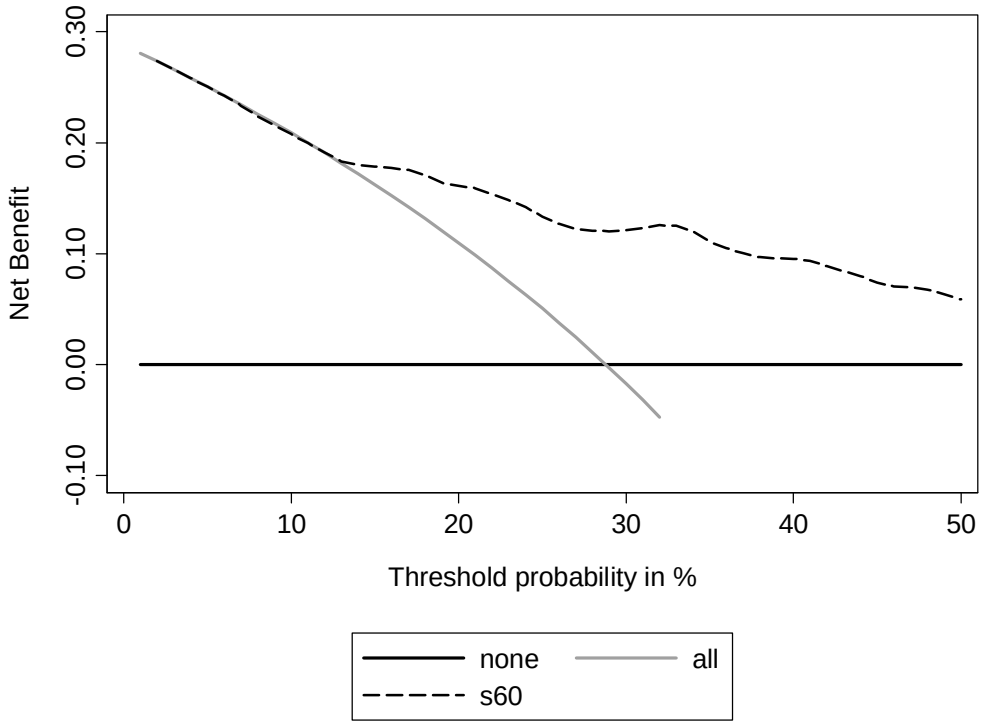
We then obtain the baseline survival probability to time t, which is stored in s0 for all observed censoring and event times. If no patient was observed at exactly time t, we can use the baseline

survival probability to the observed time closest to, but not after, time t. We'll use a timepoint of 60, which corresponds to 60 months or 5 years in this data set.

```
* the baseline survival probability to time-point 60 is stored in a local
*we take all survival times less than 60
*the lowest survival probability is the survival probability closest to 60
sum s0 if _t <= 60
local base = r(min)
```

The survival probability can then be computed as:

```
g s60 = `base' ^ exp(xb)
* now run decision curve
stdca s60, timepoint(60) smooth xstop(.5)
```



You'll notice the following warning message appear: "note: s60 was subtracted from unity so that it represents the probability of failure". This is not a problem, because s60 represents the probability of freedom from failure. We could have instead calculated the probability of failure:

```
g f60 = 1 - `base' ^ exp(xb)
stdca f60, timepoint(60) smooth xstop(.5)
```

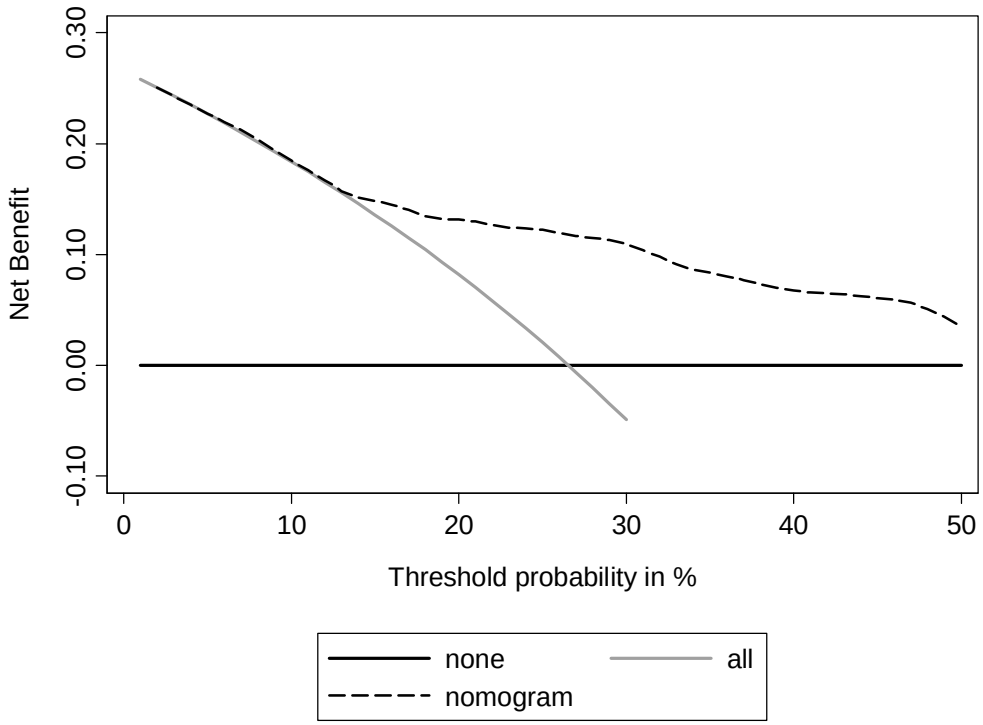
In this case, the warning message does not appear.

Bladder cancer data sets are subject to the competing risk of death. To run a competing risk analysis, we first create a failure variable that indicated patients who died before they recurred. Using the traditional approach, patients are coded 0 if they do not have an event, 1 if they have the event of interest before the competing event and 2 if they have the competing event before the event of interest.

```
g firstevent = 0
replace firstevent = 1 if recurrence==1
replace firstevent = 2 if recurrence==0 & dead==1
```

We then stset the data and run dca specifying the competing risk.

```
stset ttlastfollowrec, f(firstevent=1)
stdca nomogram, timepoint(60) cmprsk(firstevent) smooth xstop(.5)
```



The competing risk model does not change our conclusion that the statistical model would improve clinical decision making for all decisions save those for which treatment would be recommended for even very low risks, such as 2%.

The code below shows how to get both the standard decision curve and the decision curves adjusted for competing risk on one graph. It is a good example of how graphs can be drawn by saving out results from *dca* and then using the *graph* command.

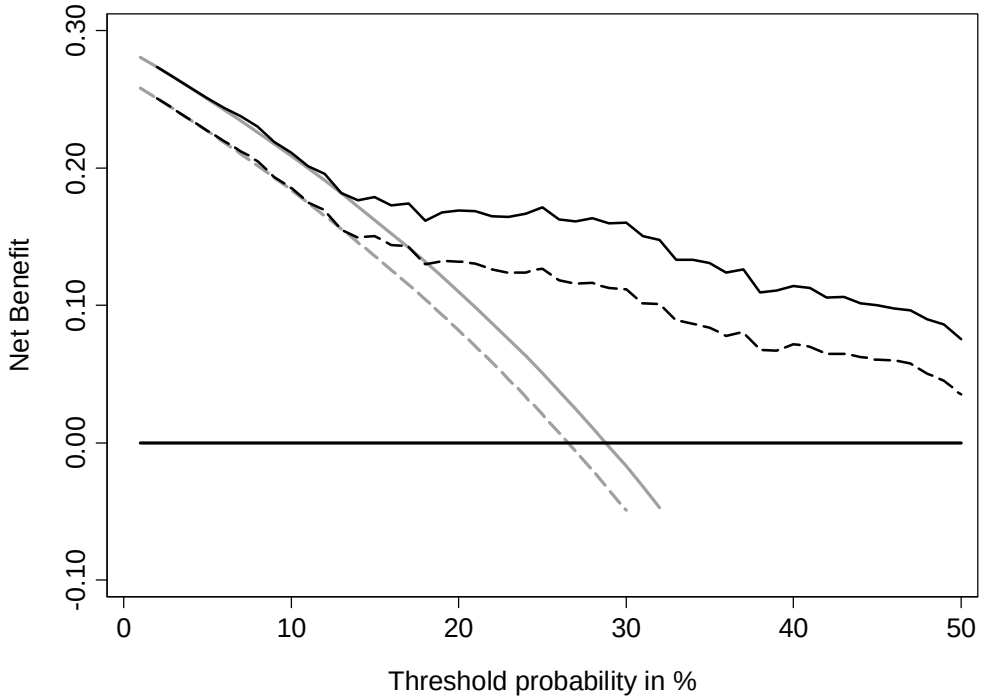
```
* start with the standard Kaplan Meier model, saving the results to a temporary file
stset ttlastfollowrec, f(recurrence)
tempfile km
stdca nomogram, timepoint(60) saving(`km') graphoff smooth  xstop(.5)

* now do the competing risk model, again saving the results to a temporary file
stset ttlastfollowrec, f(firstevent=1)
tempfile cr
stdca nomogram, timepoint(60) cmprsk(firstevent) ///
        saving(`cr') graphoff smooth xstop(.5)

*now merge the results
*start with the Kaplan Meier estimates
use `km', clear
*sort so that it can be merged later
sort threshold
*rename the variables so that we know it is the Kaplan-Meier model
rename modelp1 kmmodel
rename all kmall
tempfile kmsort
save `kmsort'

*now repeat for competing risk model
use `cr', clear
sort threshold
*do the actual merge
merge threshold using `kmsort'
rename modelp1 crmodel
rename all crall

*now create the graph
twoway(line kmall kmmodel crall crmodel none threshold, sort ///
        clpat(solid solid dash dash solid)  ///
        clwidth(medthick medium medthick medium medthick) ///
        clcolor(gs10 gs0 gs10 gs0 gs0)), ///
        scheme(s1mono) xtitle(Threshold probability in %, margin(medium)) ///
        ytitle(Net Benefit, margin(medium)) ///
        ylabel(, format(%9.2f)) legend(off)
```



Here the decision curves adjusting for competing risk are shown with dashes; the decision curves with patients censored at the time of death are shown as solid lines. Competing risk shifts all lines down, that is, estimates a lower net benefit, because censoring at the time of death overestimates the risk of death.