BayesContinuousDiagnosticTest: A program for Estimation of Disease Prevalence and the Characteristics of a Continuous Diagnostic Test and One or Two Dichotomous Diagnostic Tests

(version 3.9, September 2017)

1. Introduction

BayesContinuousDiagnosticTest is a software package for Windows to calculate marginal Bayesian posterior distributions of all unknown parameters via Gibbs sampling when data are available from a continuous diagnostic test and also possibly one or two dichotomous diagnostic tests. It is an extension to continuous tests of the methods for dichotomous tests presented in

Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard
Joseph L, Gyorkos T, Coupal L
American Journal of Epidemiology 1995; 141(3), pp. 263-272

and

Inferences for likelihood ratios in the absence of a gold standard
Joseph L, Gyorkos T
Medical Decision Making 1996; 16(4), pp. 412-417.

These papers should be read before this program is used.

Summaries of the marginal posterior densities for the prevalence of a disease, the properties of the dichotomous diagnostic test(s) [when used] (sensitivity, specificity, positive and negative predictive values and positive and negative likelihood ratios) and the properties of a continuous diagnostic test (mean score and standard deviation in positive and negative populations, sensitivity, specificity, ROC curve, positive and negative predictive values and likelihood ratios for a spectrum of cut-off values) in the absence of a gold standard diagnostic test will be output.

BayesContinuousDiagnosticTest requires that the free software packages R, WinBUGS, Perl and Microsoft Windows Script be installed (see InstallInstructions.html).

The use of this software package is straightforward, follow the directions given below.

2. How to use **BayesContinuousDiagnosticTest**

The inputs required for this program are:

• Diagnostic test data from a continuous test, used alone or together with one or two dichotomous diagnostic tests, each test applied to a group of subjects;

• Prior distributions for the disease prevalence and the sensitivity and specificity of each dichotomous diagnostic test, in the form of parameters of a beta distribution;

• Prior distributions for the continuous test scores in both positive and negative populations in the form of two normal distributions.

Continuous test score X will be modeled as a normal distribution in both positive and negative subjects, that is, it is assumed that

$X \mid \text{true status is positive} \sim N(\mu^{+}, \sigma^{2+})$
$X \mid \text{true status is negative} \sim N(\mu^{-}, \sigma^{2-})$

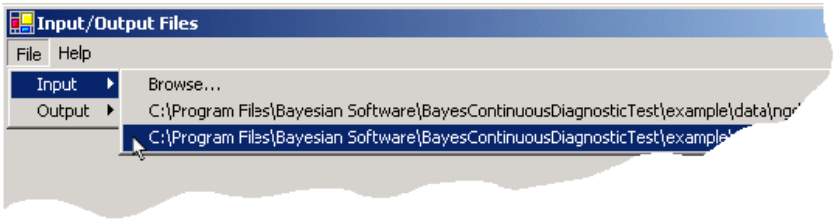
with uncertainty around means and s.d.'s expressed as a normal distribution for the means and a uniform distribution for the standard deviations, i.e.,

$$\mu^{+} \sim N(\mu_0^{+}, \sigma_\mu^{2+}) \quad \text{and} \quad \sigma^{+} \sim U(\sigma_l^{+}, \sigma_u^{+})$$
$$\mu^{-} \sim N(\mu_0^{-}, \sigma_\mu^{2-}) \quad \sigma^{-} \sim U(\sigma_l^{-}, \sigma_u^{-}).$$

where μ_0^{+} is the prior mean of the continuous test results for subjects who are truly positive for the condition under study, σ_μ^{2+} is the prior variance for μ^{+} representing how well μ^{+} is known,
and σ_l^{+} and σ_u^{+} are the lower and upper bound of the uniform prior giving the range of possible values for σ^{+}, the scores' standard deviation in positive subjects; variables superscripted with a minus sign are the corresponding values in negative subjects.

Once it is installed (see separate installation instructions), open **BayesContinuousDiagnosticTest** by clicking on the Start button on your desktop and browse to **BayesContinuousDiagnosticTest** in the Programs menu.

The initial window is used to load the data file and to select the output file location.



Data must be entered in a plain ascii text file (for example, use NotePad rather than Word) prior to being used in **BayesContinuousDiagnosticTest**.

Each column of the text file displays the results of one of the continuous or dichotomous diagnostic tests test, and the first row must consist of variable names for these tests (one word per column only: tests labels can be given to each test used in later forms). Second and subsequent rows must contain the data for all tests used for each subject, one subject's data per row. Dichotomous tests' results must be entered as 0's and 1's, 0's indicating a negative test result and 1's indicating a positive test result.

The first few lines of data file example\data\ngd+spoligo.txt (included in this package) are shown on the right.

ngd	spoligo
125.72	0
27.03	0
124.62	0
157.66	0
67.88	0
67.57	0
123.95	0
135.14	0
231.87	0
45.05	0

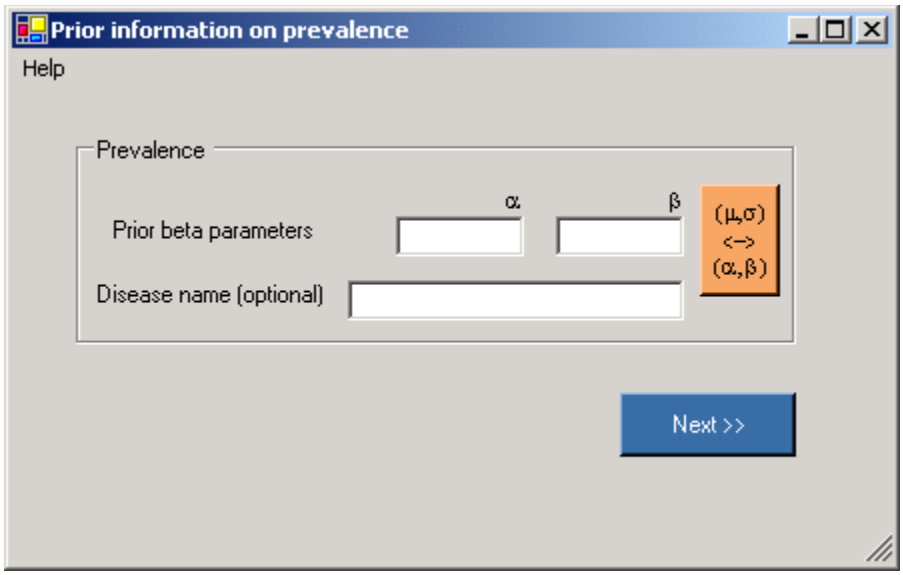
Here we have two tests, a continuous test called ngd, and a dichotomous test called spoligo. If there was a third dichotomous test, there would be a third column of data added. Data for ten subjects are displayed.

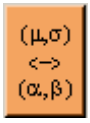
If the **Decimal symbol** is a comma on your computer (see section 3, p. 10 for details), commas in numeric values will be interpreted as a decimal symbol; that is, 125,72 will read as 125.72.

Do not use commas as digit grouping symbol, as in 123,456.78: such a value (including both a period and a comma) would cause the program to crash: one would rather enter 123456.78 for that value.

The order in which test results are entered is not important (but must be the same from line to line, of course!). **BayesContinuousDiagnosticTest** will read the data file and differentiate the dichotomous test(s) from the continuous test by the only presence of 0's and 1's in its (their) values.

The next form is used to enter your prior information on the prevalence of the disease, which is given a beta distribution with parameters (α,β), such that prior mean and variance are
$\alpha/(\alpha+\beta)$ and
$\alpha\beta/(\alpha+\beta)^2(\alpha+\beta+1)$.

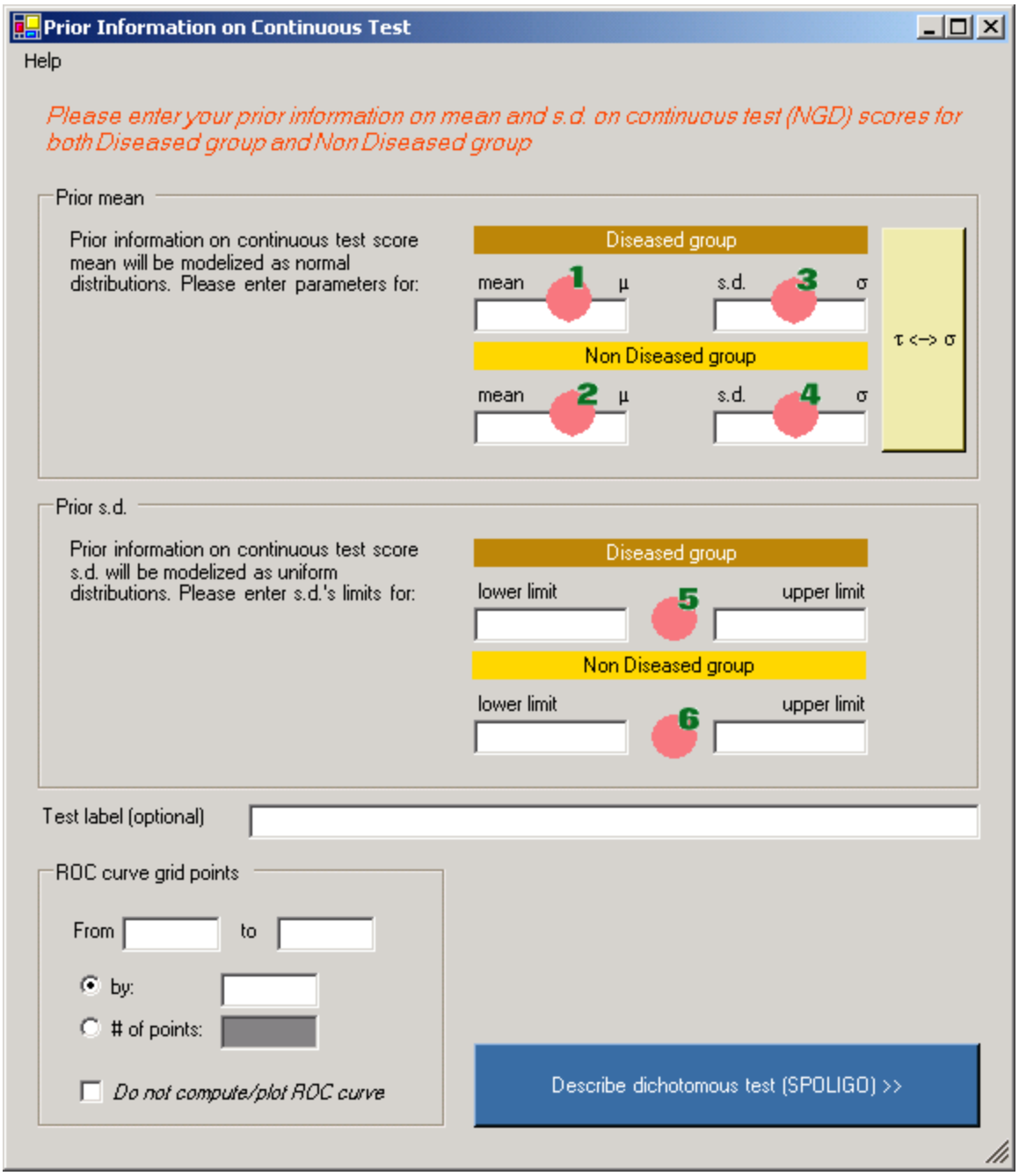




In this and other similar forms, the orange button with text (μ,σ) <–> (α,β) allows you to specify your prior distributions in terms of prior moments (μ,σ) rather than in terms of (α,β). If you choose to enter your prior information using (μ,σ), the corresponding (α,β) vector will be calculated automatically for you.

Entering a label in the *Disease name* cell will make the values from the prevalence prior entered only one click away the next time you run this program. In addition, this label will be used in all subsequent forms and in all outputs produced by this program.

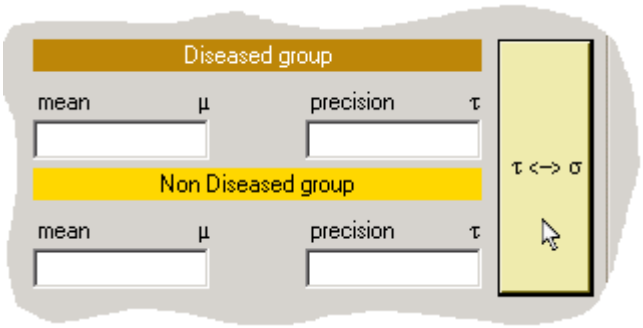
The next form (pictured on next page) will be used to enter the prior information on the continuous test score; there are more parameters to enter compared to the Prevalence or Dichotomous Tests forms, as prior separate parameters from continuous tests need to be entered for truly positive and truly negative subjects. In the picture below, the numbered dots do not appear on the original form, but are added here to facilitate the explanations for each of the corresponding cells.



As discussed earlier continuous test scores are modeled using normal distributions in both diseased and non diseased groups. Both means and s.d.'s are unknown and prior distributions are given by, in turn, a normal distribution for the mean and a uniform distribution for the standard deviation; parameters for each of these prior distributions will be entered via the two boxes labelled Prior mean and Prior s.d. on the upper half of the form.

Cells identified by the dots labelled 1 and 2 in the above picture are the prior means in both groups (using the notation introduced earlier in this section, these correspond to μ_0^+ and μ_0^-), while cells identified by the dots labelled 3 and 4 are the scale parameters for the prior means distributions (they correspond to σ_μ^+ and σ_μ^-).

Scale parameters can also be entered in terms of precision τ (where $\tau = 1/\sigma^2$, which may be preferred by some Bayesian statisticians) rather than in terms of s.d. by clicking the τ <–> σ button. Clicking it again will bring you back to the s.d. parameterization, its value being computed from the precision value entered before.



Cells surrounding the dot labelled 5 are the lower and upper prior limits for continuous test s.d. in the diseased group (they correspond to σ_l^+ and σ_u^+ in section 2.1), while cells beside dot 6 are the lower and upper limits allowed for continuous test' s.d. in the non diseased group (they correspond to σ_l^- and σ_u^- in section 2.1). These upper and lower limits represent the parameters of the uniform distributions that are used to model these standard deviation parameters.

It is important to not be confused by the parameters σ_μ^+ (σ_μ^-) and σ^+ (σ^-). The parameters σ_μ^+ and σ_μ^- represent your prior knowledge about the mean values of the continuous test result in the diseased and non diseased groups, respectively. Low values of σ_μ^+ will imply that you are confident that your choice of positive subjects' mean value is close to the truth, while larger values indicate more uncertainty about your prior value for the mean. On the other hand, σ^+ represents the degree to which the value of the continuous test varies from one subject to another within the group of disease positive subjects, and similarly for σ^- within negative subjects.

The box labelled ROC curve grid points (lower left corner of the form) allows you to define a spectrum of cut-off points (through the *From/to* cells and one of the *by/# of points* cell): **BayesContinuousDiagnosticTest** will estimate the sensitivity and specificity (at each point such defined) of the continuous test as if it were to be used as a dichotomous test, and plot the values obtained for sensitivity and specificity to draw the ROC curve. Entering 0 in the *From* cell, 200 in the *To* cell and 10 in the *by* cell will indicate the program to use the points 0,10,20,30,…,200 as cut-off points; entering 41 in the *# of points* cell (rather than using the *by* cell) will indicate to divide the 0-200 range into 40 equal length intervals, that is, to use the cut-off points 0,5,10,15,20,25,…,200.

Tick the box besides *Do not compute/plot ROC curve* if your interest is limited to the posterior distributions of the location and/or scale parameters of the test scores in diseased and non diseased groups.

Entering a label in the *Test label* cell will make this continuous test' prior information only one click away the next time you use **BayesContinuousDiagnosticTest**.

After prior parameter values for the continuous test have been entered, windows similar to the prevalence prior window will appear for the sensitivity and specificity for each dichotomous test used. Enter beta prior parameters values for these tests as you did for the prevalence.

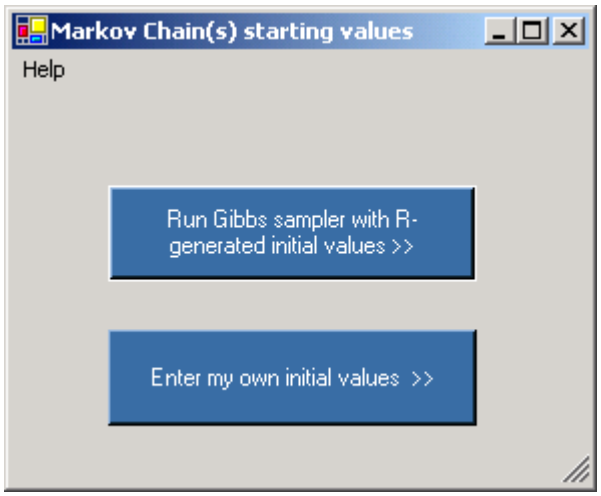
The next form is Gibbs sampler specifications form: it will allow you to control the number of burn-in iterations and the number of monitored iterations for each run of the Gibbs sampler. Burn-in iterations are iterations that are ignored when the summary statistics are calculated and are used to allow the Markov chain to converge; the history plots in the *.odc* file can be examined to assess convergence (of course, more formal convergence checks can be done, but this is beyond the scope of this document).

If you are not sure what values to select here, the default values that are filled in will in most circumstances be reasonable choices.

In general, this program should be run several times with different starting values to check for convergence. In our experience, this is especially important when the number of subjects is large or when one or two dichotomous tests are used along with the continuous test. Several chains can be run simultaneously, each with different starting values by choosing the number of runs in the menu at the right of the form.

Sensitivity runs for the prior distributions may also be important, depending on your application. See the references at the start of this article for more details about sensitivity analysis and Gibbs sampling in this application.

The next form allows you to choose whether to run the Gibbs sampler with R-generated initial values or to choose to input your own set(s) of initial values.



If you did not choose to use R-generated initial values, the next form allows you to pick your own initial values for every parameter for each of the chains you selected.

IMPORTANT: At the time that **BayesContinuousDiagnosticTest** was first released (July 2005), it was not possible to set your own seed for the random numbers generator in WinBUGS scripts (used by **BayesContinuousDiagnosticTest**): thus, answers will be identical from run to run if you always select the same initial values.

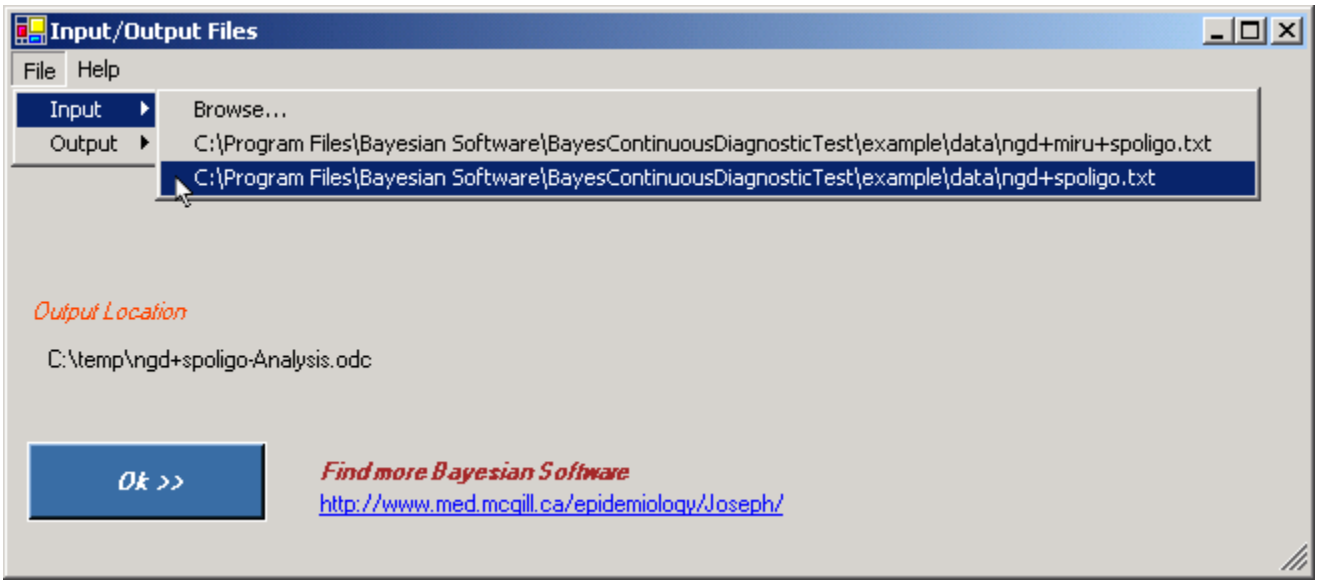
The next form will allow you to review all information entered and offer you the opportunity to change any mistyped or inaccurate information. After this last validation, **BayesContinuousDiagnosticTest** will write a WinBUGS script, run WinBUGS itself using this script (you'll see a WinBUGS' icon on your taskbar while it's running) and create a series of output files, which you will see listed in a final form (which will appear after a short delay to allow the WinBUGS program time to run).

3. An Example of Running **BayesContinuousDiagnosticTest**

To illustrate the use of this new software package, we will walk you through an example.

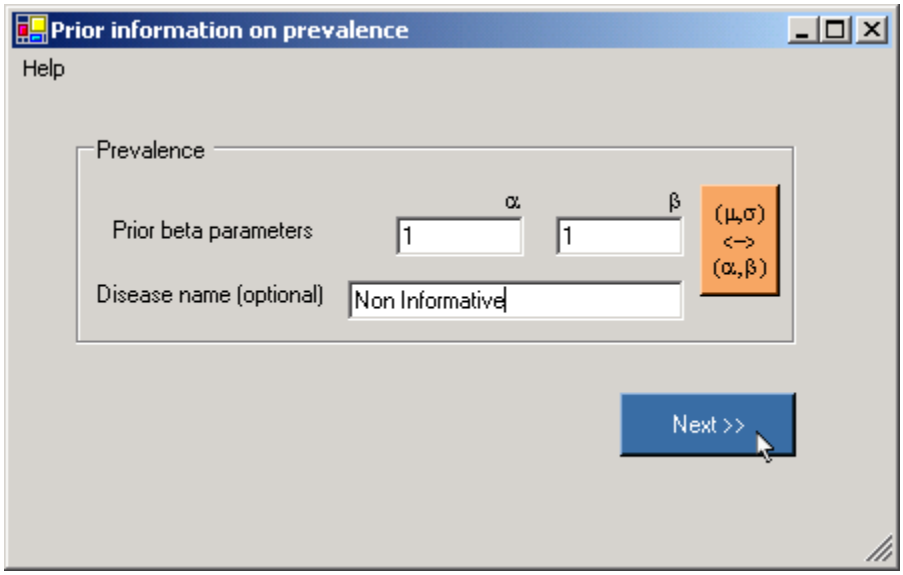
If it is the first time you are running this program, you will be prompted for the locations of your R and WinBUGS executable files shortly after opening **BayesContinuousDiagnosticTest**; make sure to indicate the most recent WinBUGS executable file found on your system (WinBUGS version 1.4 or later). If only one executable file is on your computer for each of these programs, the default locations suggested here will usually be correct.

Once this is done, the form for data entry and output file location should pop up. Select the data set where NGD (name of the continuous test for this example, this name will vary from data set to data set) is used with a dichotomous test (Spoligo); this file (ngd+spoligo.txt) is provided for illustrative purposes with this software package and should be part of the initial *File…Input>* list. Do not forget to select the output file location as well: specify a file name with the *.odc* extension [WinBUGS output extension], but note that a series of files with similar names will also be created (see details below). Click *Ok>>* when done entering both input and output file locations.



Since little prior information was available on the prevalence of the disease in this particular example, indicate a uniform prior ($\alpha=\beta=1$). You may want to enter a label for this non informative prior if you feel it is going to be used again in the future.

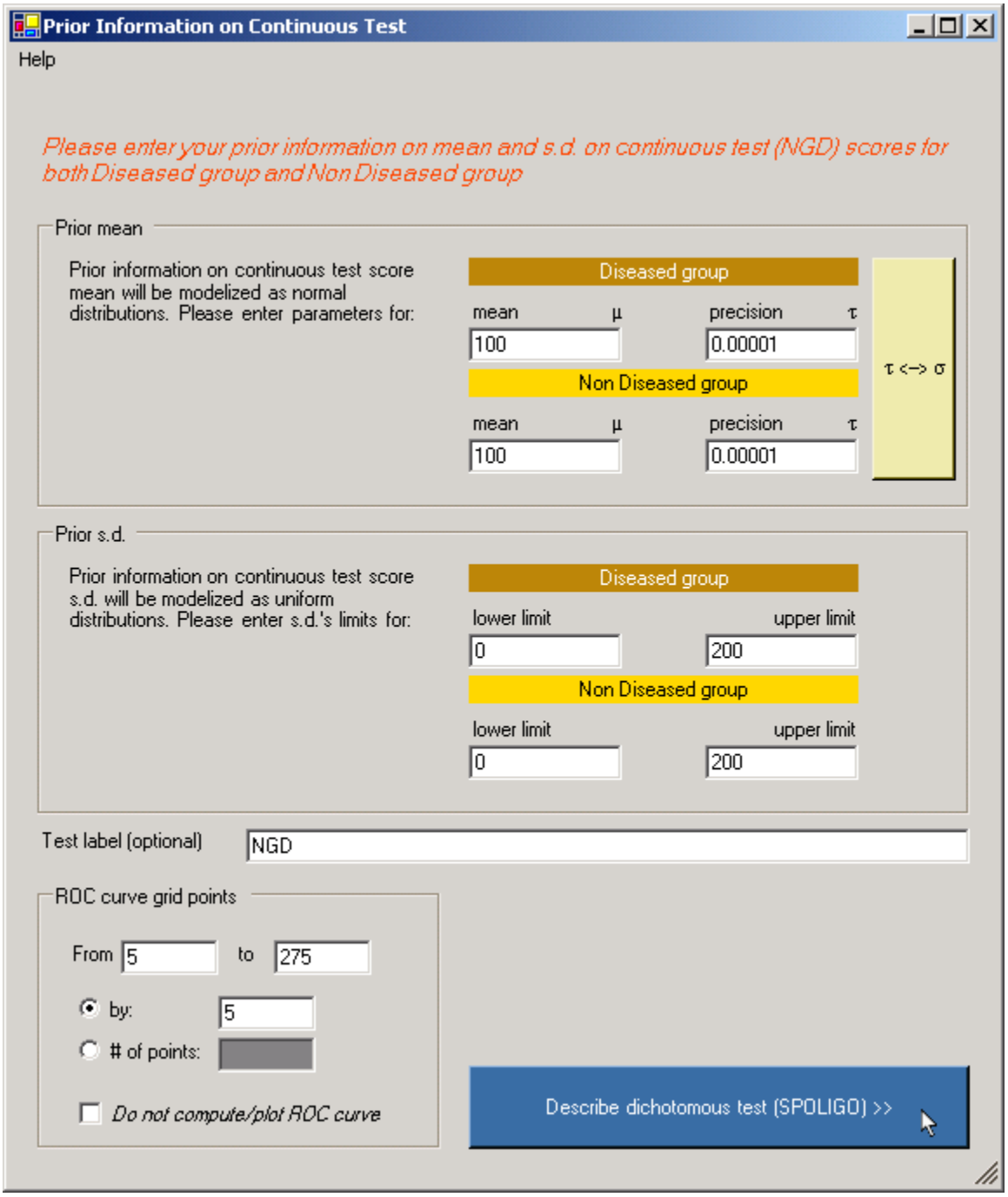
Click the *Next*>> button.



We also had only weak knowledge of the scores of NGD in positive and negative subjects; we thus decided to give relatively vague prior parameter values to its means and s.d.'s, and gave identical prior distributions for the scores in diseased and non diseased cases. When comes the time to select initial values, however, we will pick a higher initial mean for non prevalent group, as it is expected that smaller values will be observed in diseased cases. Given the range of possible values for NGD (roughly, 0-250), we will model our prior knowledge by

$$\mu^{+} \sim N(100, \tau=0.00001) \quad \text{and} \quad \sigma^{+} \sim U(0, 200)$$
$$\mu^{-} \sim N(100, \tau=0.00001) \qquad\quad \sigma^{-} \sim U(0, 200).$$

Fill in the continuous test form as shown below: do not forget to click the τ <–> σ button before entering the values for precisions.

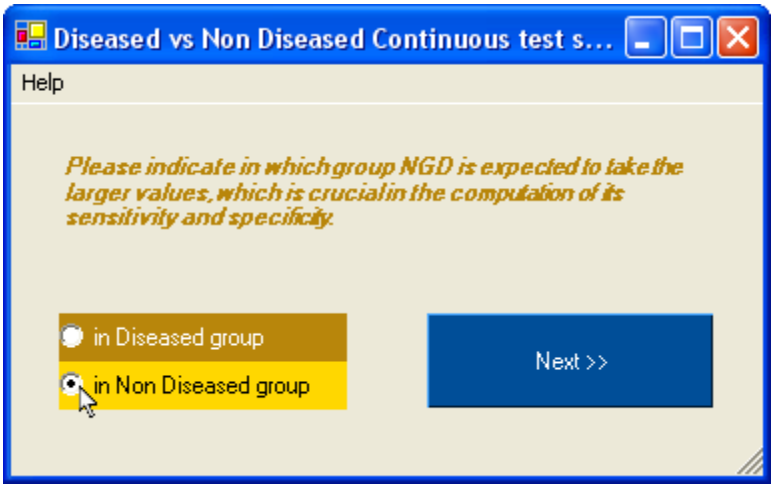


Those using a comma rather than a period as a decimal symbol should continue to use this style, assuming this concords with the setting under **Decimal symbol** in the **Customized Regional Options** form of the **Regional and Language Options** in the **Control Panel**. In this example, one would enter 0,00001 rather than 0.00001 in the precision cells.

Click on *Describe dichotomous test [SPOLIGO]>>* to continue.

When means entered in Diseased and Non Diseased prior means differ (above), their original location will serve as an indication on whether higher or lower values in the continuous diagnostic test (NGD, in this case) indicate a positive test.

When the means entered do not differ (as is the case in this example), however, the user will be prompted a form to clearly indicate the direction in which the continuous variable is expected to go. Select the *Non Diseased group* and click *Next>>*.



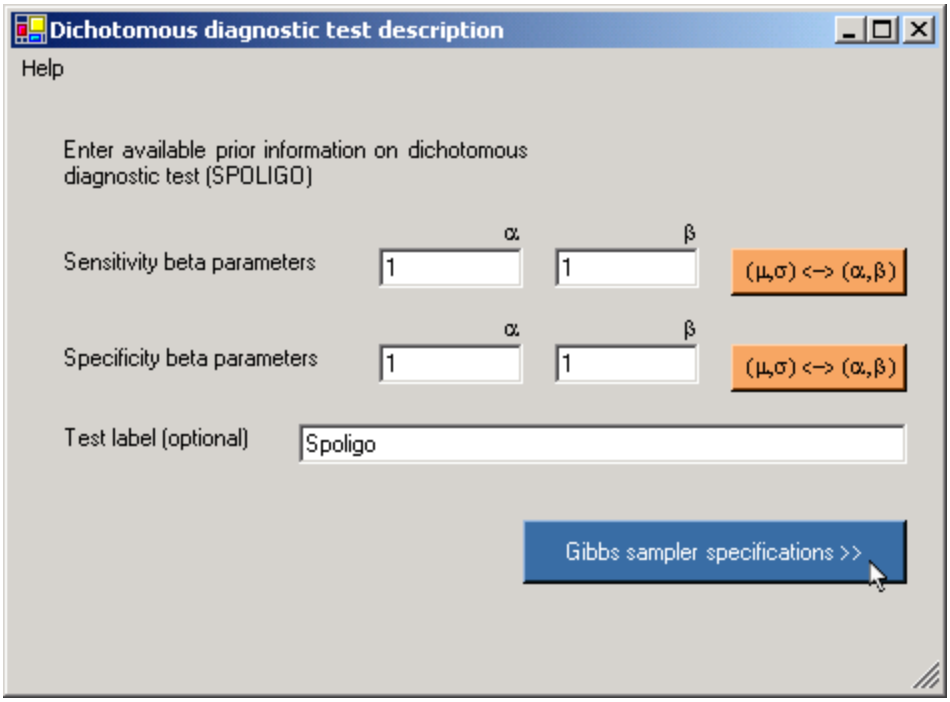
Warning:
In diagnostic testing problems, there are always two solutions that explain the data with equal probability. This duality arises since if we reverse what is considered a positive test (or the distributions of the two groups) and also let

$$prevalence(new) = 1 - prevalence(old)$$

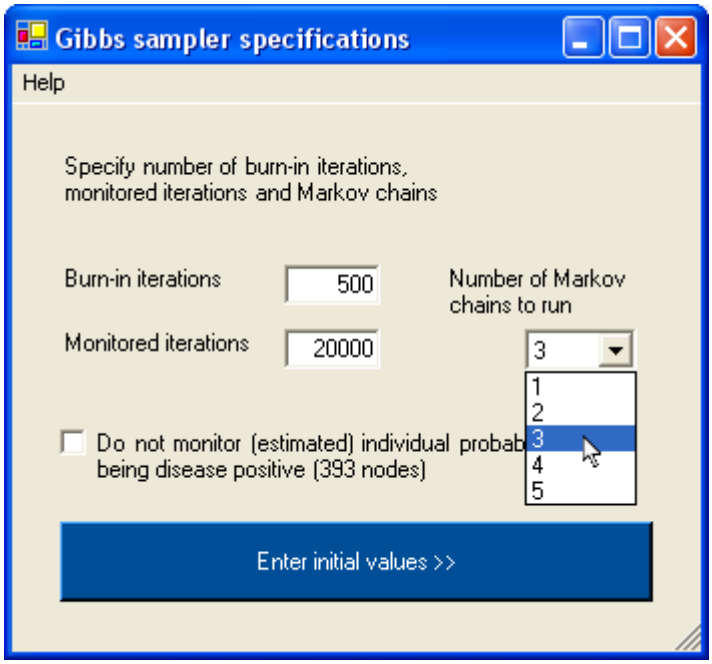
then the problems are equivalent. In practice, this might lead to an estimated prevalence of 25% when the true prevalence is (100 – 25)% = 75%.

Only limited information was available for the dichotomous Spoligo test; we thus again use uniform priors for the sensitivity and specificity of this test.

Fill in the cells as shown besides, and click *Gibbs sampler specifications>>*.



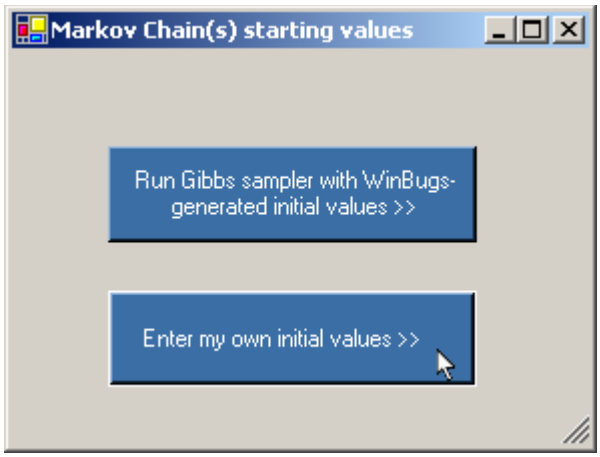
For this example, we will accept the default numbers of iterations (burn-in of 500 and monitoring of the next 20,000); also, we will run only one Markov chain (the figure to the right shows that you could run as many as five Markov chains simultaneously, but we'll stick to one in this example).



The check box in the lower half of the form allows the option to monitor the estimated individual probabilities of being disease positive or not. The advantage of not monitoring is decreased use of RAM as the program runs. The trade-off is that without monitoring these probabilities there is no plot of *Prevalence percentages* (see p.16).

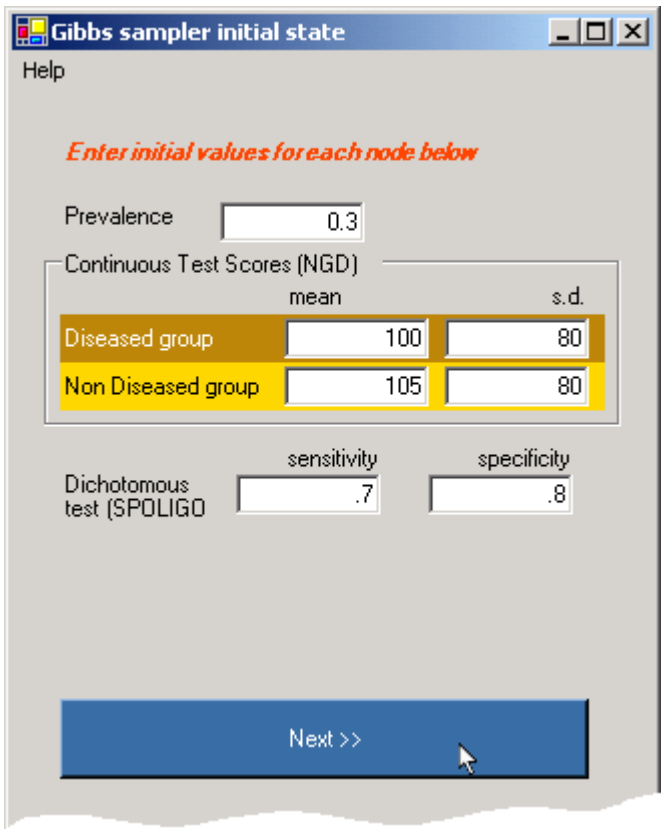
Click the *Enter initial values*>> button.

We will start the chain with our own initial values; click the *Enter my own initial values* >> button.

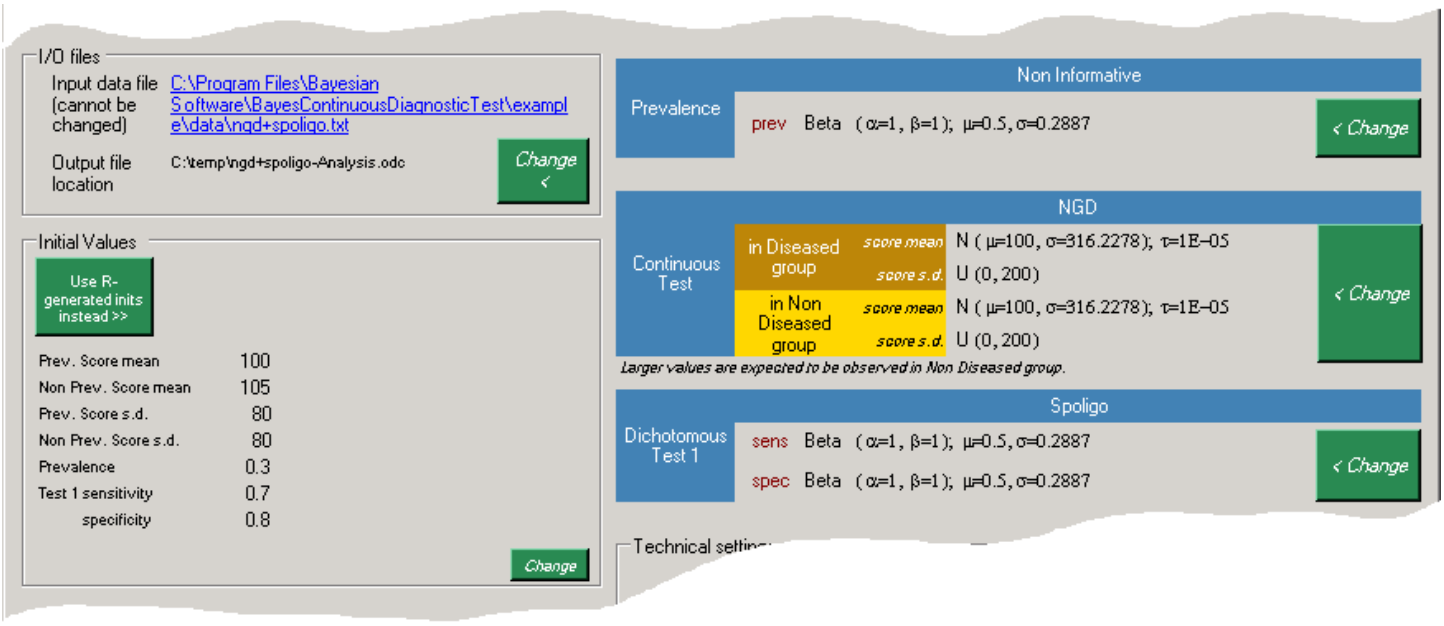


The next form allows you to pick your own initial values for every parameter for each of the chains you selected.

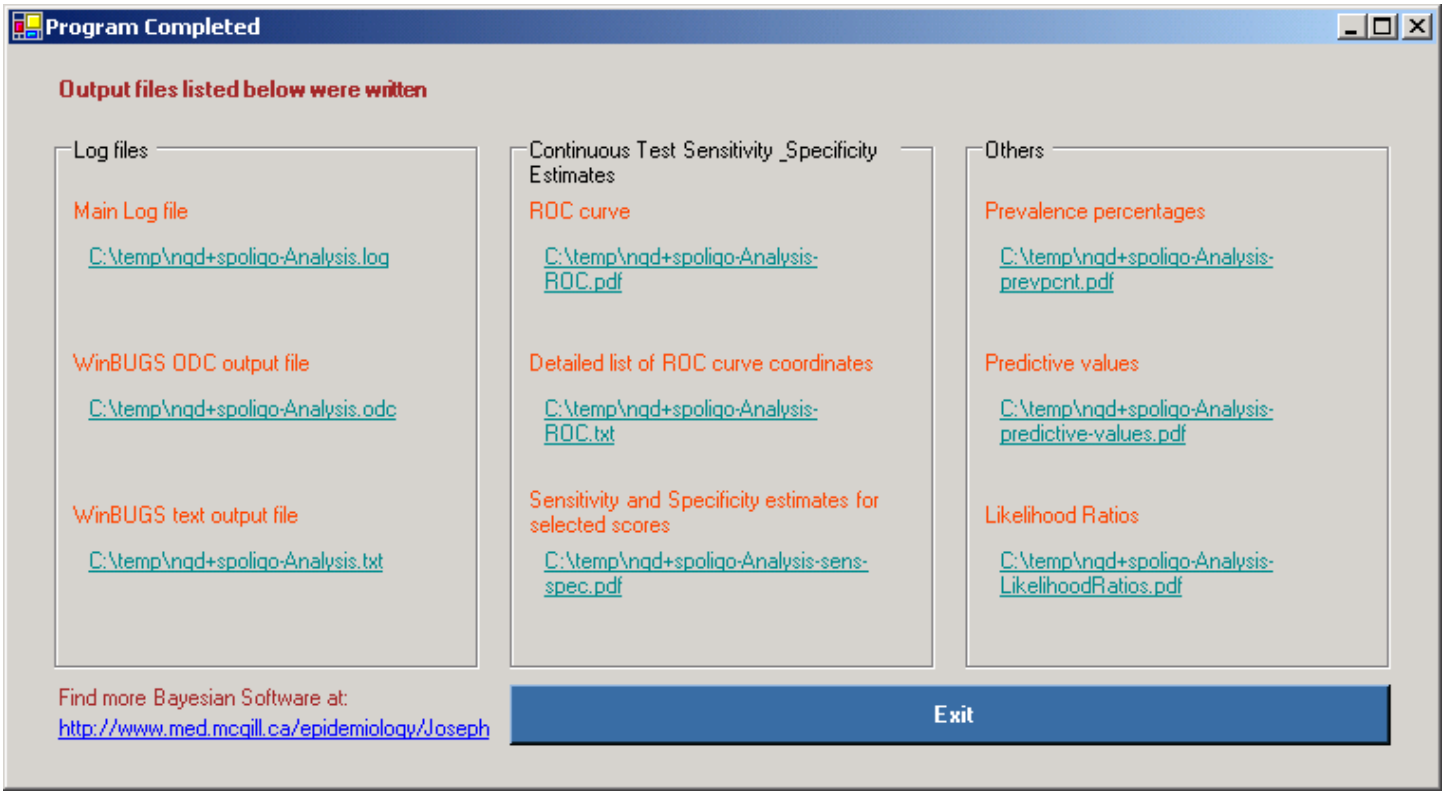
Enter the following initial values and click the *Next>> button*.



The next form summarizes all the information you have entered; please check all information and labels carefully. If you want to change any of the inputs you have entered, click the appropriate *Change* button, which will bring you back to that parameters input screen.



Click the *It's all fine. Proceed to Gibbs sampling* >> button in top right corner (not shown).
WinBUGS will then be launched: you may want to minimize the WinBUGS window while it is running (this example will typically take less than two minutes to run, but others can take longer). When done, you will be prompted by last form, giving the complete list of files written by **BayesContinuousDiagnosticTest**; note that all output files' names are augmented versions of the output file name you selected in the first form.
Click any of the output files listed if you want to view them now, or note their location to view them later. These outputs contain a wealth of information about your diagnostic tests and prevalence of the condition, based on your data and the prior parameters you used, and include both text summaries of parameter estimates as well as various graphics such as ROC curves (if that option was selected).



We reproduce below an excerpt from the Main Log file.

Main nodes statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample	
prev	0.1697	0.0351	0.001335	0.1172	0.165	0.2604	501	20000	
continuous.test.mean[1]	113.4	3.538	0.09785	106.9	113.2	121.0	501	20000	*non diseased*
continuous.test.mean[2]	26.8	3.98	0.1771	22.03	25.95	39.31	501	20000	*diseased*
continuous.test.sd[1]	49.27	2.118	0.02695	45.29	49.21	53.57	501	20000	*non diseased*
continuous.test.sd[2]	11.37	3.344	0.152	7.684	10.54	21.93	501	20000	*diseased*
likelihood.ratio.neg[1]	0.08356	0.07205	0.001062	0.002982	0.0644	0.2673	501	20000	
likelihood.ratio.pos[1]	1.648	0.1034	0.002106	1.458	1.644	1.866	501	20000	
negative.predictive.value[1]	0.983	0.01572	3.045E-4	0.9422	0.9874	0.9995	501	20000	
positive.predictive.value[1]	0.2512	0.05054	0.001938	0.1752	0.2446	0.3835	501	20000	
sensitivity[1]	0.9658	0.02916	4.375E-4	0.8925	0.9734	0.9988	501	20000	
specificity[1]	0.4124	0.03077	6.745E-4	0.3568	0.4113	0.476	501	20000	

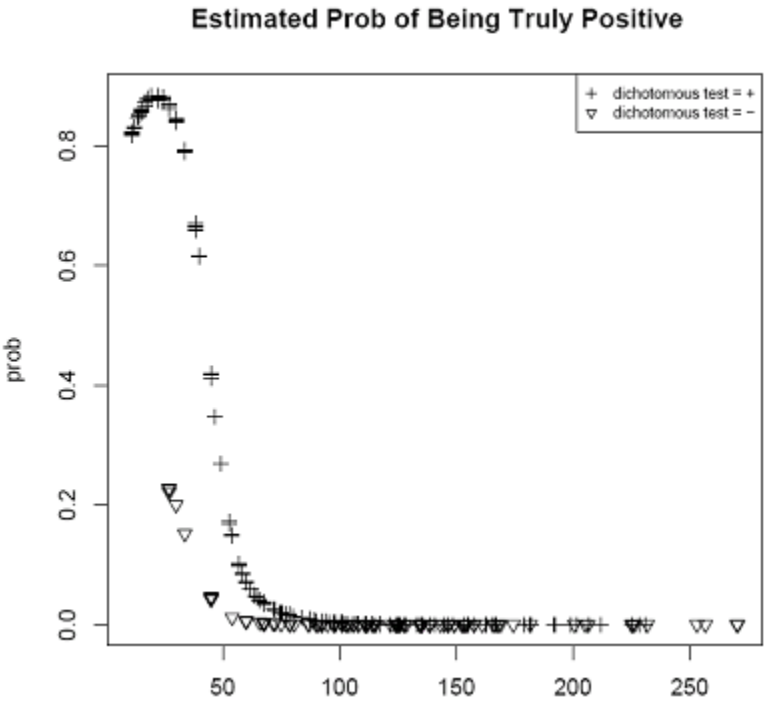
The first estimated parameter (labelled prev) gives the statistics for the prevalence of the condition under study.

The next block of parameters summarizes the characteristics of the continuous test scores among diseased subjects [2] and non diseased subjects [1] (see the diseased/non diseased labels to the right of the statistics).

A third block of statistics summarizes dichotomous tests' characteristics (when used); numbers in brackets (in this third section) refer to test numbers.

Other output files contain standard output information when dealing with diagnostic tests and should be easy to interpret for users familiar with diagnostic tests. Please note that the Likelihood Ratios plot is displayed on a log scale and that relevant WinBUGS nodes are called log.likelihood.ratio in both WinBUGS program and output files, as log values are used within and output from WinBUGS.

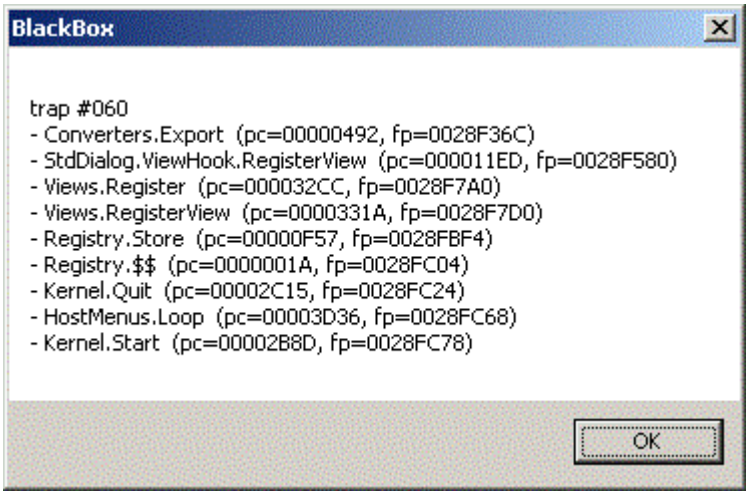
The only output file which may look unfamiliar is the *Prevalence percentages* plot, reproduced below.



The X-axis on the plot above represents the score of the continuous test while the Y-axis is the estimated probability that a subject with that test score is classified as disease positive, each point representing a subject (+'s for subjects testing positive on dichotomous test, ▽'s for negative testing subjects, as given in the upper-right corner legend). Note that the lower scoring subjects have a lower probability of being disease positive, which is in line with our earlier specifications. High scoring subjects (above 100, say) all have a very small estimated probability of being truly positive, whether they tested positive or negative on dichotomous test. On the other hand, a subject with a score of 50 on continuous test would have (roughly) an 18-20% probability of being truly positive if he also tested positive on dichotomous test, while this probability falls to 1-2% if he tested negative on dichotomous test.

4. Avoiding trap errors from permission settings on Windows 7 and Windows Vista platforms

If you are working on a Windows 7 or Windows Vista platform and have run WinBUGS before, you may have already run into the cryptic **Trap #060** error message illustrated to the right. This is due to restricted write permissions in c:\Program Files, where you may have installed WinBUGS.



WinBUGS **must** be installed in a directory where you have write permissions (e.g. C:\Users\user name \Documents) for **BayesContinuousDiagnosticTest** to run smoothly.

5. Change log

Version 1.1 (September 2005) In version 1.0, likelihood ratio for a cut-off value x was normalized, i.e., it was calculated as the ratio $f_+(x)/(f_+(x) + f_-(x))$, where $f_+(x)$ and $f_-(x)$ are the posterior continuous test probability density functions evaluated at x for the positive and negative groups respectively; it is now computed as $f_+(x)/ f_-(x)$.

Also, a minor bug fix update: version 1.0 would cause a WinBUGS trap message to appear if the path where files were saved were too long (longer than 119 characters). Version 1.1 avoids this trap by detecting troublesome long path names, and saving some files in temporary locations at WinBUGS exit, and then moving them to their proper locations after WinBUGS closes.

Version 1.2 (February 2006) A minor bug fix update: earlier versions would cause a WinBUGS trap message to appear when no dichotomous test was used.

Versions 1.3 and 1.3.1 (February 2008) A minor bug fix update: earlier versions would cause a WinBUGS trap message when one of the computed ROC curve grid points, rather than being exactly 0, was, owing to round off error, close to 0 and saved in scientific notation - which WinBUGS does not accept.

Version 1.4 (April 2008) In earlier versions, default paths to R and WinBUGS may not have been defined properly on Windows x64 platforms. This has now been corrected.

Version 1.5 (June 2008) Earlier versions accepted commas in numeric inputs, which, depending on their placement, could have led to unintended inputs being used. If a comma is found, a pop-up box now asks you to remove it, eliminating all ambiguity.

Versions 1.6, 1.6.1 and 1.6.2 (October 2008) Method for *Prevalence percentages* plot (see p. 16) changed. Before, we used the latent data for being disease positive in each iteration. We have now changed this to the estimated probability of being disease positive within each iteration, which should improve precision. Furthermore, monitoring of the latter values (and thus the plotting of *Prevalence percentages*) is made optional.

Version 2.0 (December 2008) Whether higher scores in the continuous test involved in the data analyzed indicate a positive or a negative test is crucial when estimating all parameters. In earlier versions, that information was derived from the initial values given by the user (if any) for the continuous test means in the Diseased and Non Diseased groups. This, however, seems less natural than simply deriving that information from the relative location of the prior means given to each group, which is how **BayesContinuousDiagnosticTests** now derives that information. Apart from being a more natural choice, it also allows the user to run the program with any choice of initial values, including those with unexpected locations, which was not allowed in earlier versions.

Furthermore, the choice of initial values for true status (whether a patient is – temporarily – classified as Diseased or Non Diseased) is now left to WinBUGS.

Version 2.1 (January 2009) The program now accepts either commas or periods as decimal symbols, depending on Regional Settings. Both commas and periods cannot be used at the same time, you must use the option chosen for your computer. See section 3 (p. 10) for full details.

Version 3.0 (February 2011) The Main Output file is now an .html file.

Version 3.0.1 (May 2011) A minor update that may help in correctly identifying the path to R in the initial run (especially for Windows 7 and Windows Vista users).

Versions 3.1 - 3.1.2 (December 2011) The previous default application folder (c:\Program Files) caused write permission problems for some Windows 7 and Vista users. Default application folder now changed to C:\Users\user name\Documents.

Versions 3.2 and 3.2.1 (February 2012) Minor technical problem solved from previous version.

Versions 3.3 and 3.3.1 (April 2012) We suggest a solution to prevent Trap errors for Windows 7 and Windows Vista users.

Version 3.4 (July 2012) Minor update: cmd.exe now closes automatically when program terminates.

Version 3.5 (August 2012) The path to the sub-directory where temporary files are stored was added to the Help menu of the initial form. While you can usually ignore these files, they can sometimes be helpful in troubleshooting when there are problems.

Version 3.6 (October 2012) **BayesContinuousDiagnosticTest** must limit the length of temporary paths, since it uses WinBUGS scripts, which limits file paths to a maximum of 119 characters. Longer names will cause WinBUGS to freeze. Therefore, if the default temp directory path is too long, **BayesContinuousDiagnosticTest** will ask the user to enter a path with a shorter name.

Version 3.7 (January 2015) Minor update.

Version 3.7.1 (April 2015) Minor bug fix update: a potential installation problem was solved.

Version 3.8 (January 2016) Minor update.

Version 3.8.1 (September 2016) Bug fixed from previous version.

Version 3.9 (September 2017) **BayesContinuousDiagnosticTest** now works on Windows 8 & 10. Windows 7 users do not need to reinstall or upgrade.

Questions? Comments? Please send email to: Lawrence.Joseph@McGill.ca

---

Other Bayesian software packages are available at
http://www.medicine.mcgill.ca/epidemiology/Joseph.