FasterAnalytics for Healthcare – A Clinical Case Study

Introduction
DecisionQ has developed FasterAnalytics, a unique analytical package that enables researchers, analysts, and managers to use sophisticated predictive analytics from their desktop. FasterAnalytics creates high quality, predictive models from data that enable efficient review of clinical data, real-time hypothesis testing, and rapid decisions.

FasterAnalytics uses a modeling approach called Bayesian Networks to provide a mapping of the complex relationships in data, which can then be used to make high quality predictions. Users can:

- Get an instant global view of their data.
- Understand the driving factors in the data.
- Test hypotheses in real time in our model Explorer.
- Produce reports that can be exported to other applications.
- Make determinations that can help prioritize the use of scarce research resources.

Market Overview
Clinical data analysis is instrumental to both the approval of new therapies and the provision of proper medical care. The need for clinical data analysis is currently served by a combination of services firms that analyze data, decision support products that use proprietary protocols for management of care, and traditional statistical and data analysis tools.

Value to the Customer
FasterAnalytics enables both experts and non-experts in statistics to discover and leverage knowledge from large quantities of data quickly. Examples include:

- Automatically mapping data where targets are unknown to reveal correlations.
- Discovering new relationships between variables and identifying new opportunities to improve care or reduce cost.
- Identifying potential morbidities early.
- Discovering populations that may have substantially different responses from the population at large.
- Predicting the behavior of any factor or combination of factors in the model.
- Allowing analysts to develop new models in minutes, keeping pace with shifting data.

FasterAnalytics is designed for real-time environments. Bayesian models are highly effective at identifying emerging trends that can be used to either to identify potential adverse advents or improve quality of outcomes.

Product and Technology
DecisionQ Corporation has produced a range of modules that perform data analysis, modeling, visualization, reporting, and decision optimization. FasterAnalytics modules include:

- *Discretizer.* Automatically configures the data for modeling.

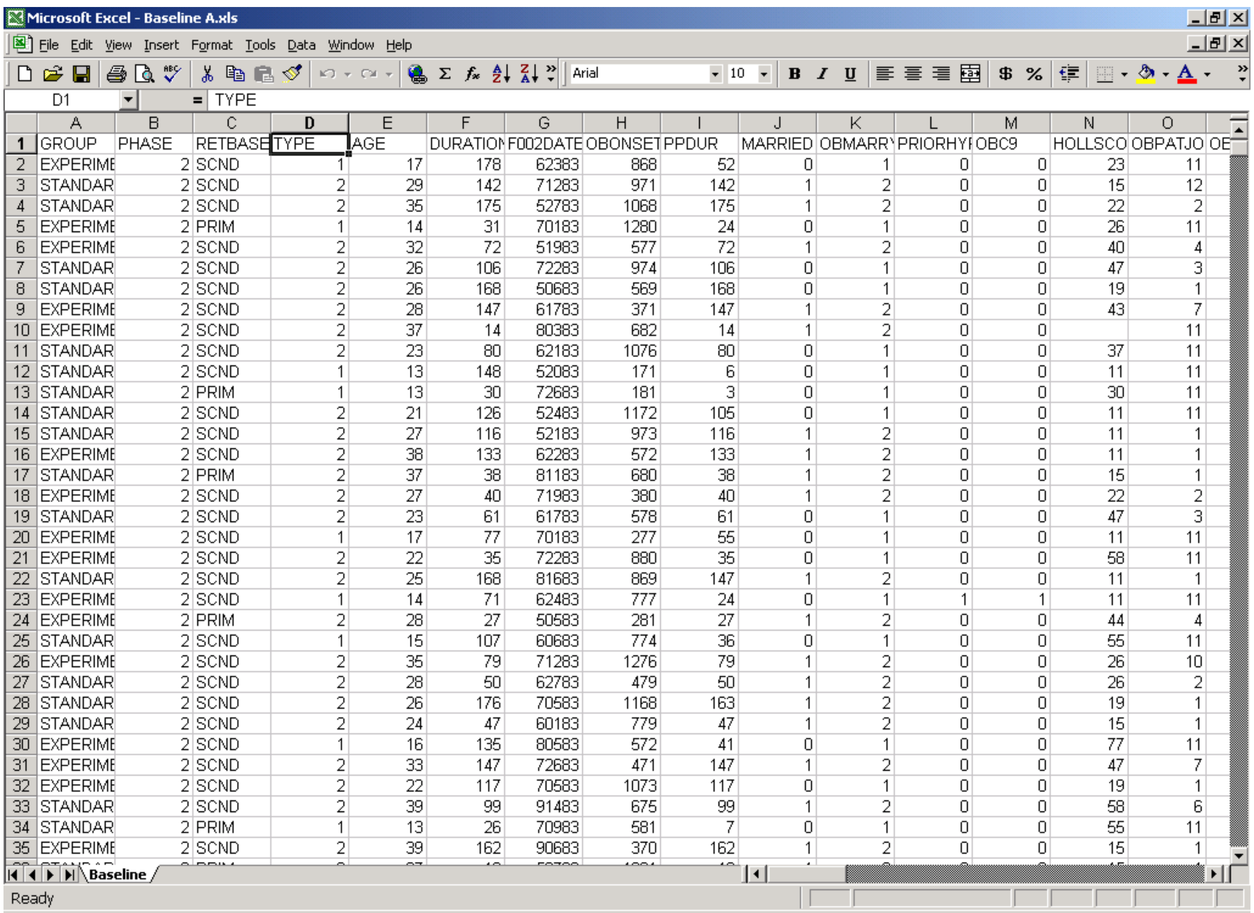
- *Modeler.* Quickly creates a visual model of the data.
- *Explorer.* Allows real-time generation and testing of hypotheses.
- *Reporter.* Extracts insights and key points for inclusion in reports and presentations.

Using the System: A Clinical Example

The following is an example of our software applied to a publicly available set of clinical data. We have used a data set comprised of the initial trial cohort for the Diabetes Control and Complications Trial, consisting of 1,441 participants reporting on 81 attributes. FasterAnalytics built the model in this example, from start to finish, in less than 30 minutes.

To build predictive models, our learning engine requires the data to be in a flat tabular format. The data can be numerical, or variable character strings. Our software can also handle missing values automatically and will either impute a value or treat missing values as a special category, at the user's discretion.

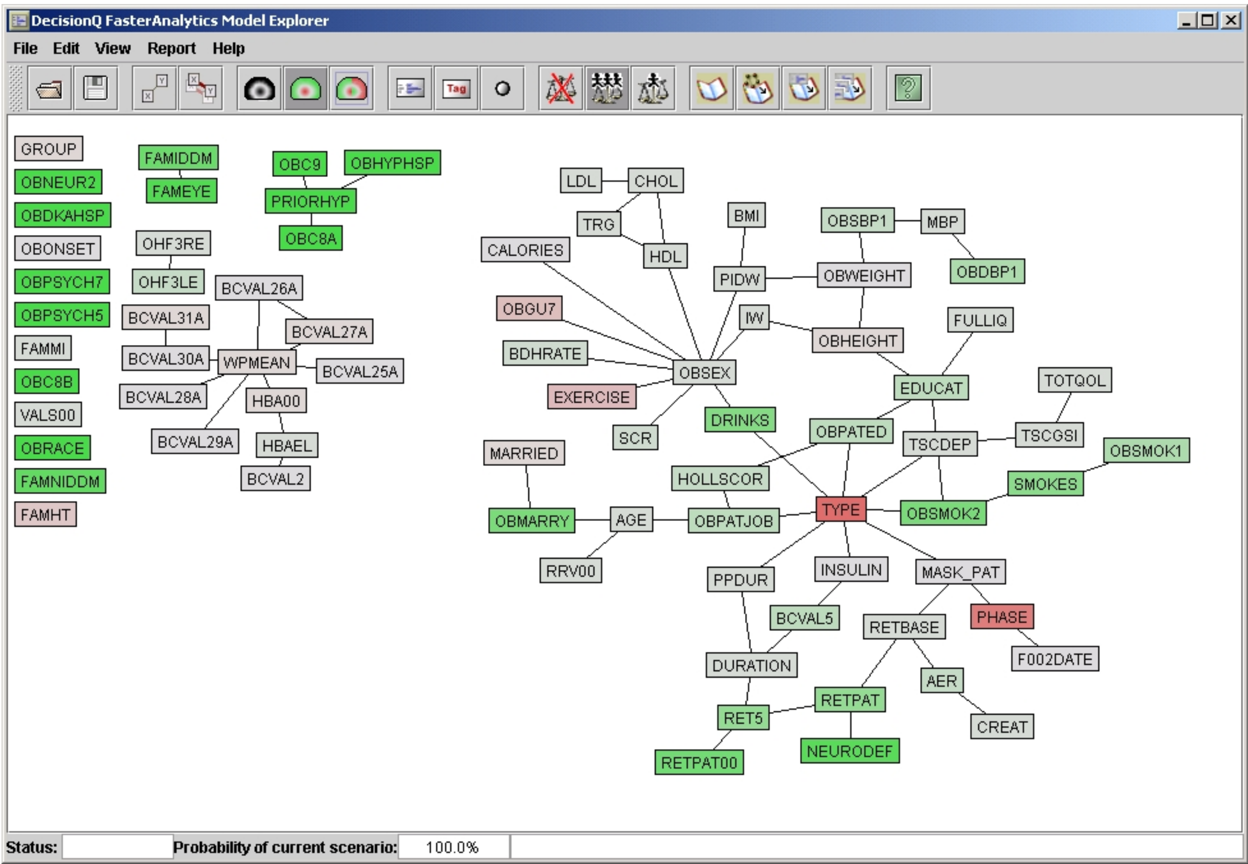
Figure 1: This example uses a data set from an Excel spreadsheet as shown below (Partial).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	GROUP	PHASE	RETBASE	TYPE	AGE	DURATION	F002DATE	OBONSET	PPDUR	MARRIED	OBMARRY	PRIORHYI	OBC9	HOLLSCO	OBPATJO
2	EXPERIME	2	SCND	1	17	178	62383	868	52	0	1	0	0	23	11
3	STANDAR	2	SCND	2	29	142	71283	971	142	1	2	0	0	15	12
4	STANDAR	2	SCND	2	35	175	52783	1068	175	1	2	0	0	22	2
5	EXPERIME	2	PRIM	1	14	31	70183	1280	24	0	1	0	0	26	11
6	EXPERIME	2	SCND	2	32	72	51983	577	72	1	2	0	0	40	4
7	STANDAR	2	SCND	2	26	106	72283	974	106	0	1	0	0	47	3
8	STANDAR	2	SCND	2	26	168	50683	569	168	0	1	0	0	19	1
9	EXPERIME	2	SCND	2	28	147	61783	371	147	1	2	0	0	43	7
10	EXPERIME	2	SCND	2	37	14	80383	682	14	1	2	0	0		11
11	STANDAR	2	SCND	2	23	80	62183	1076	80	0	1	0	0	37	11
12	STANDAR	2	SCND	1	13	148	52083	171	6	0	1	0	0	11	11
13	STANDAR	2	PRIM	1	13	30	72683	181	3	0	1	0	0	30	11
14	STANDAR	2	SCND	2	21	126	52483	1172	105	0	1	0	0	11	11
15	STANDAR	2	SCND	2	27	116	52183	973	116	1	2	0	0	11	1
16	EXPERIME	2	SCND	2	38	133	62283	572	133	1	2	0	0	11	1
17	STANDAR	2	PRIM	2	37	38	81183	680	38	1	2	0	0	15	1
18	EXPERIME	2	SCND	2	27	40	71983	380	40	1	2	0	0	22	2
19	STANDAR	2	SCND	2	23	61	61783	578	61	0	1	0	0	47	3
20	EXPERIME	2	SCND	1	17	77	70183	277	55	0	1	0	0	11	11
21	EXPERIME	2	SCND	2	22	35	72283	880	35	0	1	0	0	58	11
22	STANDAR	2	SCND	2	25	168	81683	869	147	1	2	0	0	11	1
23	EXPERIME	2	SCND	1	14	71	62483	777	24	0	1	1	1	11	11
24	EXPERIME	2	PRIM	2	28	27	50583	281	27	1	2	0	0	44	4
25	STANDAR	2	SCND	1	15	107	60683	774	36	0	1	0	0	55	11
26	EXPERIME	2	SCND	2	35	79	71283	1276	79	1	2	0	0	26	10
27	STANDAR	2	SCND	2	28	50	62783	479	50	1	2	0	0	26	2
28	STANDAR	2	SCND	2	26	176	70583	1168	163	1	2	0	0	19	1
29	STANDAR	2	SCND	2	24	47	60183	779	47	1	2	0	0	15	1
30	EXPERIME	2	SCND	1	16	135	80583	572	41	0	1	0	0	77	11
31	EXPERIME	2	SCND	2	33	147	72683	471	147	1	2	0	0	47	7
32	EXPERIME	2	SCND	2	22	117	70583	1073	117	0	1	0	0	19	1
33	STANDAR	2	SCND	2	39	99	91483	675	99	1	2	0	0	58	6
34	STANDAR	2	PRIM	1	13	26	70983	581	7	0	1	0	0	55	11
35	EXPERIME	2	SCND	2	39	162	90683	370	162	1	2	0	0	15	1

Having selected the data, the fully automated process continues until a full model is presented, or the user can stop each part of the process to manually change parameters. The software begins by categorizing the data and 'binning' in accordance with the default settings. The data is then passed seamlessly to the Modeler for automated model development. Once the software has mapped the complex relationships in the data a model is presented in the Explorer.

Figure 2: Base case model of the data presented in Explorer



The display illustrates conditional dependence between variables and the pathways existing in the cohort model. Notice that the network has multiple branches, and that the data is interrelated in a "web", one of the strengths of multivariate Bayesian networks. Once the model is created, the user has the ability to move and create more space between nodes for easier viewing of the whole model.

In the example below, we examine how selecting the Type I or Type II diabetic population affects the profile of the trial cohort. We begin by selecting our target variable, Type II. The thick border indicates that this is the target selected, and its color red indicates that we are interested in analyzing how other variables behave when the target is Type II diabetics (instead of Type I). When we toggle the value of the TYPE node between Type I and Type II, we notice the colors of other nodes change to express positive or negative correlations with this node. (See Figure 3 and Figure 4). The tint of color indicates the degree of positive or negative correlation.

Figure 3: TYPE variable set to Type II

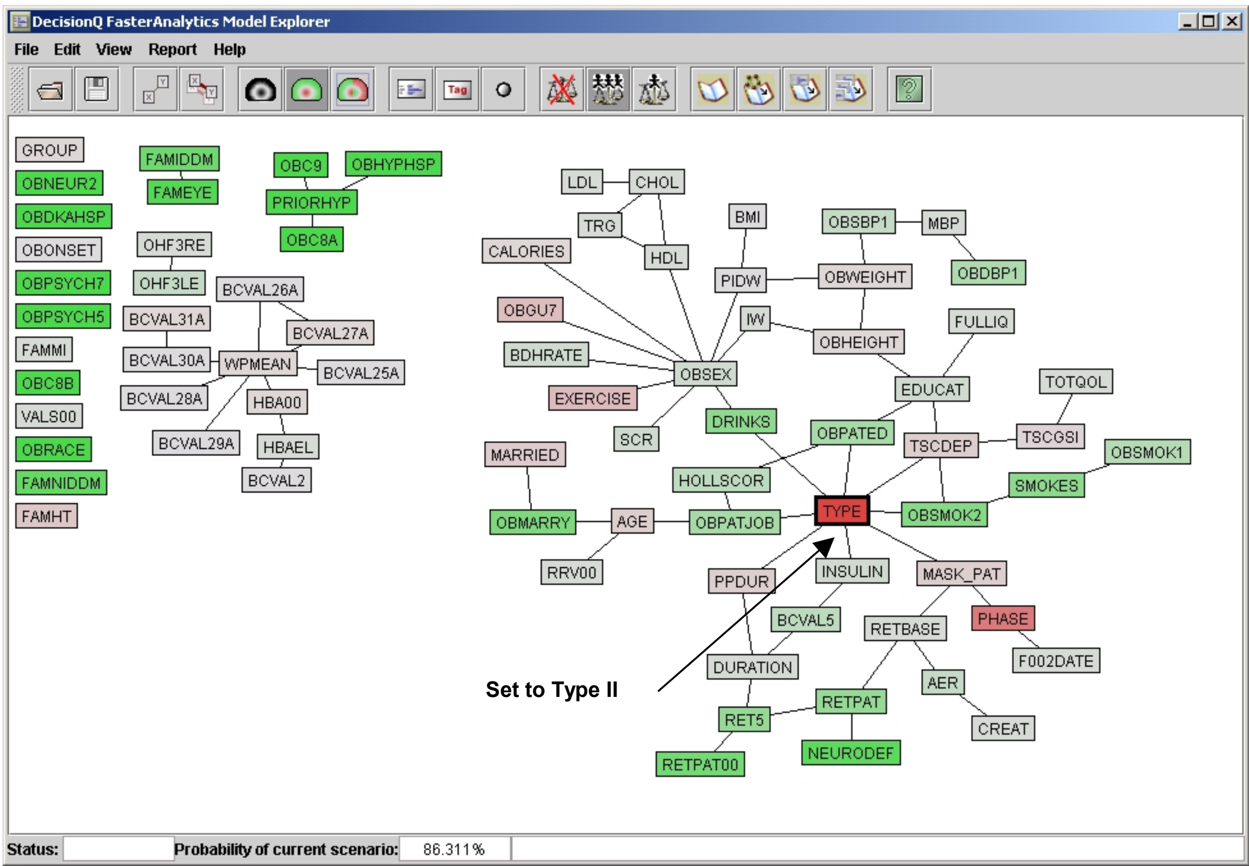
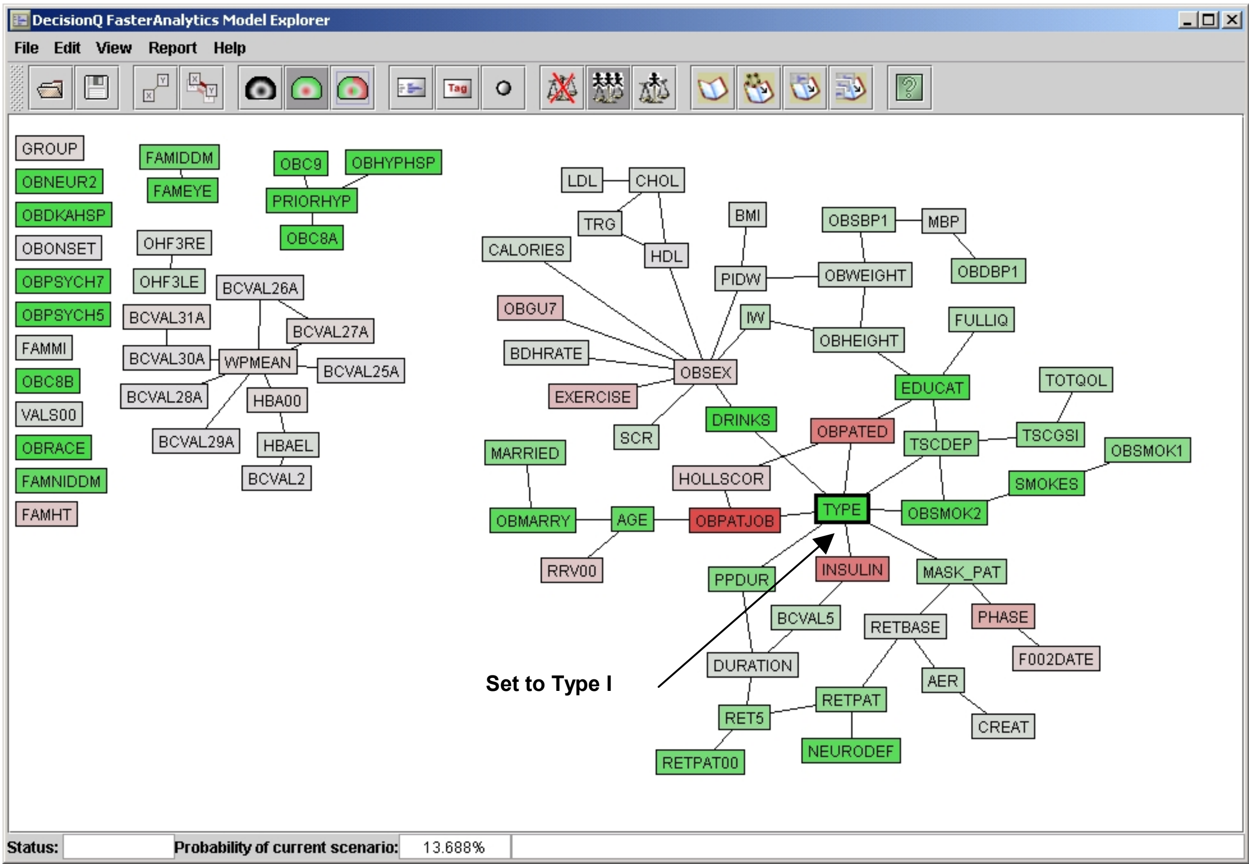


Figure 4: TYPE variable set to Type I



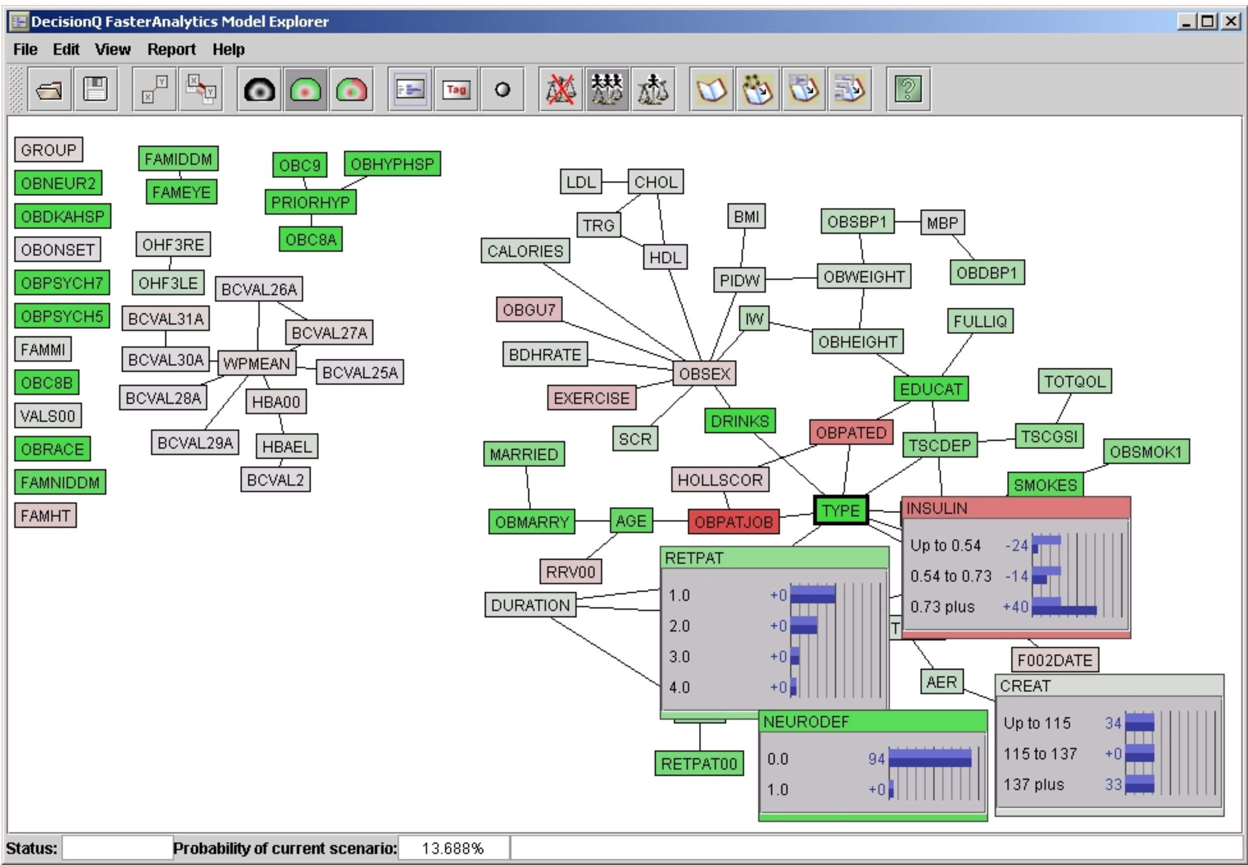
Compare the two models in Figure 3 and 4 above with the base level in Figure 1. While the Type II population is not particularly different from the overall cohort, the Type I population has a dramatically different profile. The coloring in the graphical model shows the change in population profile quickly and effectively.

It is also possible to select two or more variables simultaneously. The extent to which weight also affects the population profile can be studied in conjunction with other factors such as gender.

Each clinical and laboratory factor can be expanded using the ‘View’ menu or icons to show quantitative information about the relationships between these variables. The population data is displayed as “cases” with bars that represent the marginal probability distribution of each case.

Suppose that we are interested in examining how Serum Creatinine and Insulin levels affect Retinopathy and Neurologic Deficit levels in the Type I diabetic population. We first select these nodes and click “Graph” to display the cases within these nodes. This can be done for as many variables as we may choose.

Figure 5: Quantitative information about INSULIN, CREATININE, RETINOPATHY, and NEUROLOGIC DEFICIT



If we wish to test hypotheses, we can modify any node and see how our hypothesis affects the model. Notice how information flows through the network.

When we inquire regarding the effect of lower levels of insulin and serum creatinine, we observe very clearly that there are beneficial effects of keeping these substances low in the bloodstream as demonstrated by improvement in retinopathy and neurologic deficit. When the levels of insulin and creatinine are raised however, we do not observe an equally negative effect. We can see this graphically in Figures 6a and 6b below.

Figure 6a: Low INSULIN and CREATININE and their effect on diabetic RETINOPATHY

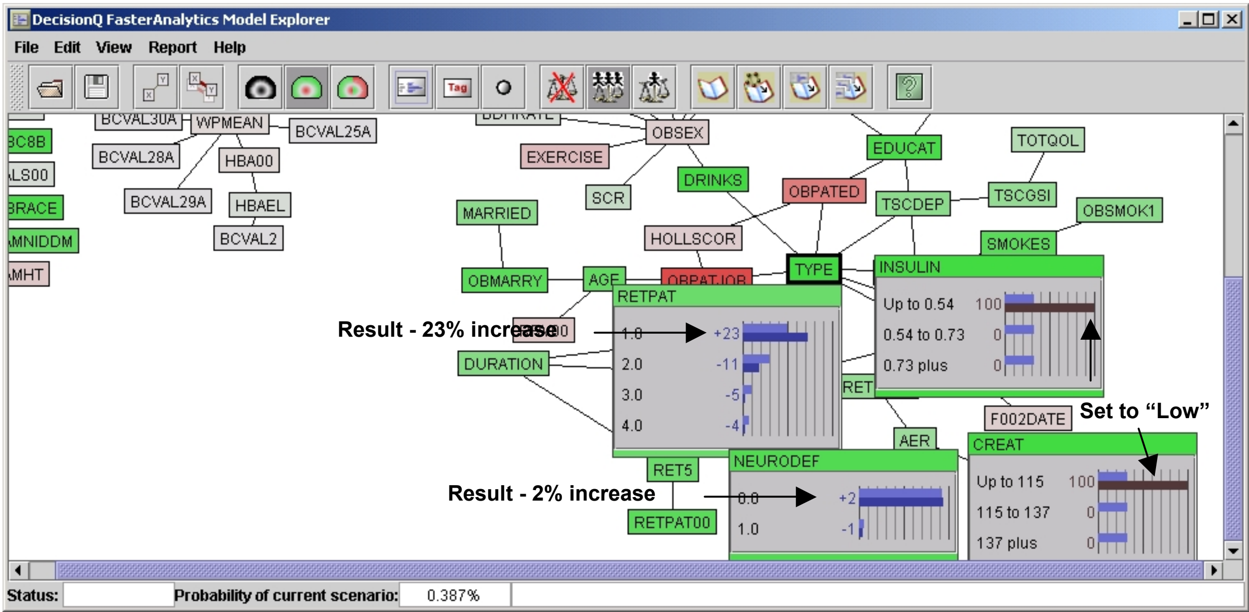
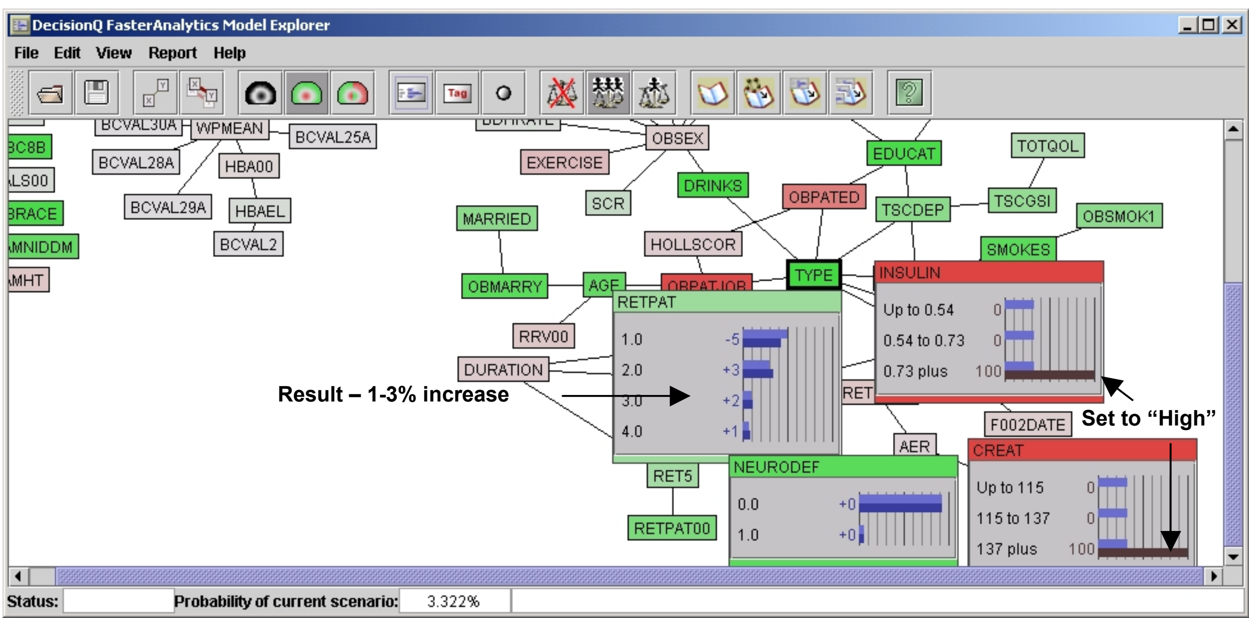
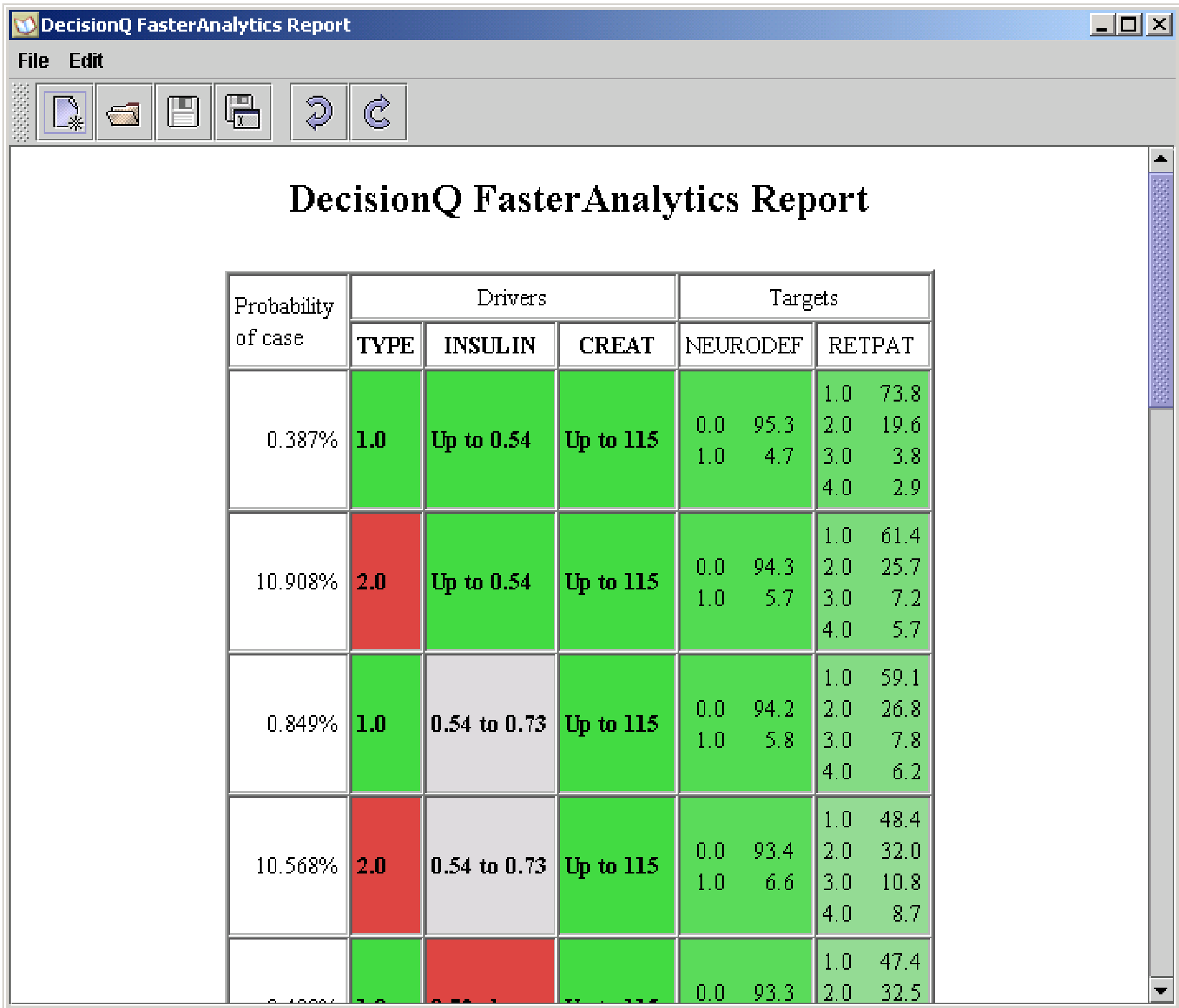


Figure 6b: High INSULIN and CREATININE and their effect on diabetic RETINOPATHY



The Report module can be used to create a report that will show the conditional probabilities (or predicted likelihood) of any target variables, given the expression of any independent variable(s). Any part of the model visualization can be pasted into Reporter and then transferred into other applications. Figure 7 shows a sample report.

Figure 7: A sample report listing the probabilities of retinopathy and neurological deficit given creatinine and insulin levels



DecisionQ FasterAnalytics Report

File Edit

DecisionQ FasterAnalytics Report

Probability of case	Drivers			Targets	
	TYPE	INSULIN	CREAT	NEURODEF	RETPAT
0.387%	1.0	Up to 0.54	Up to 115	0.0 95.3 1.0 4.7	1.0 73.8 2.0 19.6 3.0 3.8 4.0 2.9
10.908%	2.0	Up to 0.54	Up to 115	0.0 94.3 1.0 5.7	1.0 61.4 2.0 25.7 3.0 7.2 4.0 5.7
0.849%	1.0	0.54 to 0.73	Up to 115	0.0 94.2 1.0 5.8	1.0 59.1 2.0 26.8 3.0 7.8 4.0 6.2
10.568%	2.0	0.54 to 0.73	Up to 115	0.0 93.4 1.0 6.6	1.0 48.4 2.0 32.0 3.0 10.8 4.0 8.7
[illegible]	[illegible]	[illegible]	[illegible]	0.0 93.3	1.0 47.4 2.0 32.5

DecisionQ sells predictive modeling software and complementary professional services. Alternatively, components from FasterAnalytics can be integrated into third party applications as part of broad data management and analysis platform.

If you have any further questions or would like to schedule a more detailed demonstration in person or over the web, please contact us.

DecisionQ Corporation
3726 Connecticut Avenue, Suite 519
Washington, DC 20008
www.decisionq.com
Phone: 415-254-7996
Fax : 415-276-6356
Email: info@decisionq.com