



GORE™ Medical Membrane Technologies
 Your Experienced Partner for
 High-Performance Membranes for Medical Devices

Find:

In:

[Advanced Search](#)

- [Find Products & Suppliers](#)
- [Suppliers Directories](#)
- [Product & Service News](#)
- [Web Site Gallery](#)
- [News & Research](#)
- [Events](#)
- [News](#)
- [Research](#)
- [Magazines](#)
- [Medical Device & Diagnostic Industry](#)
- [Medical Product Manufacturing News](#)
- [IVD Technology](#)
- [Pharmaceutical & Medical Packaging News](#)
- [European Medical Device Manufacturer](#)
- [Medical Electronics Manufacturing](#)
- [MX: For Medtech Executives](#)
- [Tradeshows](#)
- [All Shows](#)
- [Medical Design & Manufacturing](#)
- [Packaging](#)
- [Plastics](#)
- [Automation](#)
- [Design & Manufacturing](#)
- [Career Center](#)
- [Search Jobs](#)
- [Post Jobs](#)
- [Job Seeker Login](#)
- [Register](#)
- [About This Site](#)
- [Contact Us](#)
- [Site Guide](#)
- [Awards Program](#)
- [How to Advertise](#)
- [Register](#)
- [Login](#)



Originally Published [IVDT](#) April 2006

SOFTWARE DEVELOPMENT

Interpreting diagnostic assays by means of statistical modeling

Applying multivariate probabilistic networks to pathology data in an automated system has good predictive value.

Howard Robin, John S. Eberhardt III, Rick Gaertner, Jennifer Kam, and Mike Armstrong

The typical pathology laboratory processes vast amounts of patient data daily. New immunologic techniques, and advances in molecular biology and pharmacogenomics, have led to a proliferation of biologic markers, diagnostic assays, and novel biotherapeutics that has increased both the quantity and quality of laboratory information. These developments have made the therapeutic decision-making process more complex.

However, applying dynamic statistical modeling to laboratory information may allow the physician to make more-effective use of available clinical information.^{1,2} This article assesses the use of a Bayesian modeling approach to combine known clinical and pathologic information in a mathematical framework in order to calculate pre- and posttest probability of chronic kidney disease (CKD) in the hospital emergency department. Bayesian analysis of complex data sets has historically been time-consuming and fraught with the potential for errors. New machine learning software can automate much of this process and produce a consistent result. The objective of this investigation was to evaluate one such software package, DecisionQ FasterAnalytics from **DecisionQ Corp.** (Kentfield, CA). In particular, the goal was to determine the quality, accuracy, and utility of the Bayesian models it produces.

The Investigation and Methods

Early stages of CKD are difficult to detect with current laboratory assays.^{3,4} If disease signs are detected early, effective treatment can halt or slow the progression toward end-stage renal disease.⁵ But studies have indicated that many patients are first diagnosed late in the course of their disease.⁶ While a number of studies have evaluated the use of new biomarkers for detection of early-stage renal disease, those tests are typically limited by analytic measurement, stage of disease, and the presence of other morbid conditions.⁷⁻⁹



Figure 1. [\(click to enlarge\)](#) A probabilistic (Bayesian) model of renal function in a hospital emergency department population.

For the study reported here, the investigators applied a Bayesian machine learning tool to a data set of emergency-room patients with the goal of calculating pre- and posttest probability of renal insufficiency.

Data Collection. Between June 2003 and August 2004, at the Sharp Hospital emergency departments in San Diego, 740 consecutive patients who had a routine brain natriuretic peptide (BNP) level ordered by the department physician were automatically enrolled in the study. The study used routine clinical and laboratory data; therefore, enrollment did not require consent and was approved by the Sharp institutional review board committee. An order for a BNP level triggered the analysis of N-terminal prohormone BNP (NT-proBNP) and the data required for the calculation of the estimated glomerular filtration rate (eGFR) using the modification of diet in renal disease (MDRD) formula. The MDRD formula is based on blood urea nitrogen (BUN), serum creatinine, serum albumin, age, race, and sex.

When the subjects were discharged from the hospital, whether directly from the emergency department or following hospitalization, the final discharge ICD-9 codes (from the International Classification of Diseases, 9th rev.) were obtained by the investigators from the hospital's medical records department. The investigators evaluated the following disease diagnoses: congestive heart failure (CHF), renal disease, acute myocardial infarction (AMI), chronic obstructive pulmonary disease (COPD), pneumonia, diabetes, and hypertension (HI BP).

When data collection was complete, the investigators inspected the final data set for errors, discrepancies, and significant patterns of missing data and found it to be suitable for modeling. They classified certain variables further for binning prior to modeling; in this, they followed established clinical practice and the scientific literature (see Table I).

Statistical Analysis. Data analysis involved the use of a Bayesian belief network. A Bayesian network encodes the joint probability distribution of all the variables in the domain by building a network of conditional probabilities. It uses conditional independence assumptions to make the representation tractable. The directed networks incorporate parent-child relationships between nodes. In a network, a node is independent of its nondescendants given its parents. The Bayesian networks discussed here were constructed with the FasterAnalytics software.

The software-generated Bayesian networks use machine learning to calculate prior

Variable	Value	Probability
Age	65	0.15
Sex	Male	0.55
BUN	12	0.25
Serum Creatinine	1.2	0.30
Serum Albumin	4.0	0.10
BNP	100	0.20
Diagnosis	Renal Disease	0.10
Diagnosis	CHF	0.15
Diagnosis	AMI	0.05
Diagnosis	COPD	0.05
Diagnosis	Pneumonia	0.05
Diagnosis	Diabetes	0.05
Diagnosis	Hypertension	0.10

probabilities and the structure of the network. Prior probabilities are derived from the data to be modeled through calculation of a distribution of discrete states, or by using equal-area binning in the case of continuous variables. The software derives the network structure by means of a heuristic search method that generates hypothetical models with different conditional independence assumptions.

The heuristic search takes advantage of two proprietary innovations. The first is a data caching and query system developed to allow an order of magnitude more data to be analyzed than previously possible. This enables users to consider more-complex problems in less time on inexpensive computing hardware. The second innovation uses highly efficient searches, adding flexibility to the heuristic search. Besides the greater speed and efficiency this feature brings to Bayesian modeling, the model quality score it produces is generally 1–5% better than that achieved with a standard heuristic algorithm.¹²

MDRD Variable States	Training Set				Mean
Normalized Description	1	2	4	5	
Normal	95.22	94.83	95.73	95.99	95.92
Mildly decreased	54.30	54.53	53.23	53.35	53.76
Moderately decreased	61.43	64.73	61.07	61.88	62.28
Severely decreased	20.86	19.93	21.97	20.88	20.66
Normal/acute	62.75	58.43	64.23	64.88	64.51
Other	91.28	92.48	91.29	92.58	92.12
Overall values	Mean: 66.34				66.34
	Standard Dev: 10.73				10.73

Table II. (click to enlarge) Cross-validation data for MDRD eGFR for the five training sets in the study.

All values are percentages.

probability distribution of the data.

The network was validated by means of a train-and-test cross-validation methodology. This methodology involves selecting an independent test set at random and holding it apart from the rest of the data for training the model. The test set used represented 20% of the total data set, and the training set represented 80% of the total data set for a twofold cross-validation. The researchers repeated this twofold validation five times.

Once the model was constructed from the training set, the investigators input the test set into it, generating a case-specific prediction for each patient record for variables of interest. They then used the test set predictions to calculate predicted values (PVs) and receiver-operator-characteristic (ROC) curves for each model. The ROC curve was calculated by comparing the PV for each variable to the known value in the test set on a case-specific basis. It was then used to calculate the area under the curve (AUC), a metric of overall model quality that indicates the degree of trade-off in the model between sensitivity and specificity. PVs were calculated for each outcome of a given variable. They provide the probability that a given outcome and given threshold are a true positive.

Results of Data Modeling

Modeling the laboratory data using the FasterAnalytics software, the investigators demonstrated a relationship between eGFR (using the MDRD formula) and cystatin C, creatinine, gender, diagnosis of CKD, diagnosis of CHF, and blood urea nitrogen BUN.^{13,14}

MDRD Variable States	Training Set				Mean
Normalized Description	1	2	4	5	
Normal	95.22	94.83	95.73	95.99	95.92
Mildly decreased	54.30	54.53	53.23	53.35	53.76
Moderately decreased	61.43	64.73	61.07	61.88	62.28
Severely decreased	20.86	19.93	21.97	20.88	20.66
Normal/acute	62.75	58.43	64.23	64.88	64.51
Other	91.28	92.48	91.29	92.58	92.12
Overall values	Mean: 66.34				66.34
	Standard Dev: 10.73				10.73

Table III. (click to enlarge) Predictive values for MDRD eGFR for the five training sets in the study. All numbers are percentages.

The agreement between cystatin C classified according to the Mayo Clinic modification scale and eGFR using the MDRD is evident in Figure 1.^{15,16} The model confirms the relationship between the various cystatin C classification approaches, age, and race.¹⁷ In addition, it demonstrates the direct relationship between the diagnosis of CKD and HI BP and diabetes.¹⁸ It is interesting that the ratio of NT-proBNP to BNP is conditionally related to increases in the blood urea nitrogen. Figure 1 illustrates the structure of the model and gives the reference ranges of the variables. Each modeled variable is represented as a node in the network. The histograms represent the probability distribution of outcome, with the percent probability of outcome noted in the label next to each bar.

The investigators statistically cross-validated the structure of the model using hold-out data from the five randomized train-and-test sets. They calculated ROC curves for the variable eGFR for each test set and calculated the AUC and PV for each variable state. The mean value of eGFR for the overall data set was 52.7 ml/min/1.73 m² with a 95% confidence interval of 2.4 ml/min/1.73 m². The mean value of eGFR for training sets 1, 2, 4, and 5 ranged from 52.0 to 53.0 ml/min/1.73 m², while set 3 had a mean eGFR of 57.5 ml/min/1.73 m². The 95% confidence intervals were 2.5–2.6 ml/min/1.73 m² for training sets 1, 2, 4, and 5, and were 2.3 ml/min/1.73 m² for set 3. The AUC ranged from 80.9 to 98.2% with a mean of 91.5%, and the PV of the model ranged from 25.0 to 100.0% with a mean of 66.3% (see Table II).

Table I. (click to enlarge) Patient-data variables as classified by binning methodology for Bayesian analysis using FasterAnalytics software.

The software promotes the network with the most likely model given the data available.

Models are scored according to minimum description length, a system that provides a measure of the quality of a model. This scoring technique trades off good fit for reduced model complexity. Goodness of fit is the likelihood of the data given the model, while model complexity is the amount of information needed to store the model. Minimum description length scoring is asymptotically equivalent to Bayesian scoring, also known as the Bayesian information criterion.

The software evaluated in the study uses several algorithms for handling missing data. For the investigation, the chosen algorithm ignored the parts of each record that was missing information. To prevent overfitting of the model to the data, the researchers used two approaches. First was the application of a complexity penalty that required the addition of any new parameter to result in an equal or additional amount of predictive power. Second, a Dirichlet prior was used during modeling to build an expected error into the prior



Figure 2. (click to enlarge) Calculation of expected probabilities when normal renal function is assumed.

PVs of normal renal function, moderately decreased renal function, and severely decreased renal function, while the PV of renal failure declined and the PV of mildly decreased renal failure declined slightly (see Table IV).

In addition to illustrating relationships, the model encodes joint probabilities. From an assumption of normal eGFR, for example, the expected probability of CHF, creatinine, and cystatin can be calculated (see Figure 2). The dark-red bar in the histogram represents evidence that has been set, and the dark-blue bars represent the expected probability of outcome relative to the reference population represented by the light-blue bars. The labels indicate the percentage-point change in probability of expected outcome relative to the reference population. In a patient with normal renal function, there is an expected 20-percentage-point decrease in the probability of a diagnosis of CHF, a 52-point increase in the likelihood of being in the low end of the range for creatinine, and a 32-point increase in the likelihood of being at the bottom of the cystatin range.

On the other end of the spectrum, renal failure can be designated as the assumption and the expected probabilities calculated (see Figure 3). In a patient with renal failure, there is an expected 6-percentage-point increase in the probability of a diagnosis of CHF, a 49-point increase in the likelihood of being at the top of the range for creatinine, and a 77-point increase in the likelihood of being at the top of the cystatin range.

The same models can be used to predict outcomes. For example, the investigators set evidence to assume a female patient with diagnoses of HI BP and CHF (see Figure 4). This patient can be seen to have a 14-percentage-point increase in the likelihood of moderately decreased renal function (52% versus 38% for the reference), and a 3-point increase in the likelihood of severely decreased renal function (15% versus 12%). This model calculates the most likely MDRD outcome as moderately decreased renal function, with a probability of 52%.

The information can also be displayed in an inference table that provides an expected distribution for the variable of interest for each combination of contributing factors, as well as an expected frequency for each case.

Calculations of the expected distribution of eGFR given cystatin C and HI BP have been performed (see Table V).

Discussion

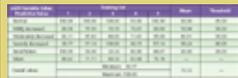


Table IV. (click to enlarge) PVs from Tables II and III as calculated using modified thresholds. All values are percentages.

The poor survival rate for patients with late-stage CKD, as well as the success in preventing the disease's progression when detected early, makes the early diagnosis of renal disease important. Unfortunately, the current markers for renal disease are unreliable for differentiating mild and moderate cases of renal dysfunction from cases of severely decreased function.

Using a probabilistic (Bayesian) model, researchers can calculate, from multiple data elements, the probability of a patient having mild to moderate CKD. The software employed in this study demonstrated that laboratory measurements of cystatin C, serum creatinine, and blood urea nitrogen, when combined with clinical information including age, sex, and diagnoses of CHF or HI BP, can improve physicians' ability to detect early renal disease. These laboratory and clinical attributes may then be used to calculate a patient-specific pretest probability of renal disease, identifying patients who require further clinical and laboratory evaluation.

The model is appropriately validated using ROC curves for all levels of renal function, with mean AUC running from 89.1% for severely decreased renal function to 95.3% for renal failure (see Table II). High AUCs (90.0%) have been achieved with cystatin C alone in patients with mild to moderate renal dysfunction; however, as Table II reveals, this is equivalent to the AUC achieved by the model (90.6–90.8%) in the same population.¹⁹ The strong AUCs for all levels of renal function, and the particularly strong PVs for the population with moderately decreased renal function, indicate that the model is potentially valuable for the detection of patients in the early stages of CKD, a clinical area in which evaluation tools currently available are limited.

The use of predictive models for research is well established. Now it is beginning to gain acceptance in clinical medicine. Algorithms for the early detection of CKD have been developed, and studies of new biomarkers for the detection of early renal disease have been reported.^{20–22} A broad framework that supports the use of multiple markers and clinical parameters has the potential to provide a more universal method for early detection.

The Bayesian model described here allows a specific framework to be developed for a given patient population, and it supports decision making based on multiple parameters and partial clinical information. The model can be tailored to the population of a specific facility, can be easily and dynamically updated, and can provide both rules of thumb and case-specific prediction. The output of the model is a probabilistic risk score that enables the clinician to assess the relative risk of early-stage CKD for a particular patient, given the local population. Testing and treatment resources then may be allocated optimally.

The FasterAnalytics graphical user interface allows the clinician to calculate pre- and posttest probability of disease and risk relative to the reference population. Construction of the model is operationally simple, and validation involves established statistical methods. Clinicians who are familiar with basic computational programs and basic statistics can construct and use similar models that are based on their domain of expertise.

The study discussed in this article does incorporate certain limitations into the data that were modeled, of course.

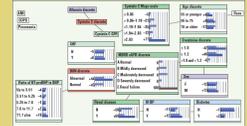


Figure 3. (click to enlarge) Calculation of expected probabilities when renal failure is assumed.

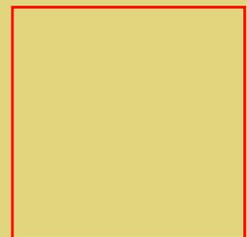


Figure 4. (click to enlarge) Expected probabilities of renal dysfunction in a patient assumed to have hypertension and congestive heart failure.

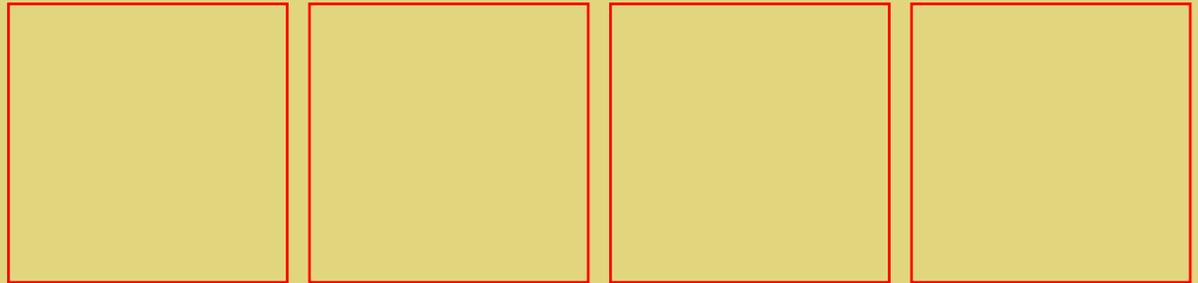
The use of ICD-9 codes rather than chart review could lead to diagnostic biases. The use of the MDRD formula rather than a quantitative marker examined in serum and urine samples from the patient also limits the clinical utility of this marker, as the model has been tested and validated using an estimate of glomerular filtration based upon creatinine. However, the promising results observed in model cross-validation indicate that a revised model using a true measure of GFR should produce a model of high clinical utility in the identification of patients with early CKD.

Table V. ([click to enlarge](#)) Expected distribution of eGFR given cystatin C and heart failure. Green = low-end distribution skew (i.e., normal kidney function). Red = high-end distribution skew (i.e., renal failure).

In addition to data limitations, there are certain limitations inherent in predictive, and specifically Bayesian, modeling. The model is constructed using a machine learning algorithm that develops the most likely model given the available data. This does not necessarily mean that the model is true or representative of the broader population. Further, as a probabilistic tool, the model provides an estimate of the likelihood of outcome. Cross-validation statistics do supply an indication of the expected error of that estimate; however, Bayesian networks and machine learning cannot be assigned a confidence level the way a traditional hypothesis test can. While these models are very useful in stratifying patients by risk, it is important that they be used in the context of other clinical information and the experience and judgment of the clinician.

Conclusion

Standard laboratory data were found to be useful in successfully creating a Bayesian model to assist in the screening of CKD in hospital emergency department patients with signs and symptoms of heart failure. A group of additional renal markers were identified that can be used in a predictive model to quantify risk of disease in populations that have historically been difficult to diagnose. The addition of a predicted probability of disease to an intuitive graphical report improves the diagnostic utility of laboratory information for the clinician.



(R to L) **Howard Robin, MD**, is chairman of the department of pathology and medical director of the clinical laboratory at Sharp Memorial Hospital (San Diego). He is also medical director of the office of continuing medical education at Sharp HealthCare (San Diego). **John S. Eberhardt III** is a founder and executive vice president of DecisionQ Corp. (Kentfield, CA). **Rick Gaertner** is the supervisor of clinical chemistry for Sharp HealthCare. **Jennifer Kam** (not pictured) is research analyst for the Sharp Memorial Hospital clinical laboratory. **Mike Armstrong** is administrative system director of laboratory services at Sharp HealthCare. The authors can be reached at howard.robin@sharp.com, john.eberhardt@decisionq.com, rick.gaertner@sharp.com, jenny.kam@sharp.com, and mike.armstrong@sharp.com, respectively.

References

01. GH Lyman and L Balducci, "Overestimation of Test Effects in Clinical Judgment," *Journal of Cancer Education* 8, no. 4 (1993): 297–307.
02. H Kataoka and T Sugiura, "The Ideal Form of Laboratory Information Management," *Rinsho Byori* 53, no. 1 (2005): 39–46.
03. AK Bello, E Nwankwo, and AM El Nahas, "Prevention of Chronic Kidney Disease: A Global Challenge," *Kidney International Supplement*, no. 98 (2005): S11–S17.
04. G Gambaro et al., "Silent Chronic Kidney Disease Epidemic Seen from Europe: Designing Strategies for Clinical Management of the Early Stages," *Journal of Nephrology* 18, no. 2 (2005): 123–135.
05. MD Wavamunno and DC Harris, "The Need for Early Nephrology Referral," *Kidney International Supplement*, no. 94 (2005): S128–S132.
06. KS Kinchen et al., "The Timing of Specialist Evaluation in Chronic Kidney Disease and Mortality," *Annals of Internal Medicine* 137, no. 6 (2002): 479–486.
07. A Ahlstrom et al., "Evolution and Predictive Power of Serum Cystatin C in Acute Renal Failure," *Clinical Nephrology* 62, no. 5 (2004): 344–350.
08. AG Christensson et al., "Serum Cystatin C Advantageous Compared with Serum Creatinine in the Detection of Mild but Not Severe Diabetic Nephropathy," *Journal of Internal Medicine* 256, no. 6 (2004): 510–518.

09. K Tamba et al., "Prospective Evaluation of Renal Function by Serum Cystatin-C: Comparison with Three Other Parameters of Glomerular Filtration Rate," *Nippon Jinzo Gakkai Shi* 43, no. 8 (2001): 646–650.
10. Timothy Larson, pers. comm.
11. A Larsson et al., "Calculation of Glomerular Filtration Rate Expressed in mL/min from Plasma Cystatin C Values in mg/L," *Scandinavian Journal of Clinical and Laboratory Investigation* 64, no. 1 (2004): 25–30.
12. J Moraleda, "New Algorithms, Data Structures, and User Interfaces for Machine Learning of Large Datasets with Applications" (PhD diss., Stanford University, 2003).
13. K Nitta et al., "Serum Cystatin C Concentration as a Marker of Glomerular Filtration Rate in Patients with Various Renal Diseases," *Internal Medicine* 41, no. 11 (2002): 931–935.
14. A Christensson et al., "Serum Cystatin C Is a More Sensitive and More Accurate Marker of Glomerular Filtration Rate Than Enzymatic Measurements of Creatinine in Renal Transplantation," *Nephron Physiology* 94, no. 2 (2003): 19–27.
15. E Wasen et al., "Estimation of Glomerular Filtration Rate in the Elderly: A Comparison of Creatinine-Based Formulae with Serum Cystatin C," *Journal of Internal Medicine* 256, no. 1 (2004): 70–78.
16. M Hertlova et al., "Cystatin C in Estimates of Glomerular Filtration in Patients with Renal Disease—Initial Experience," *Vnitřní Lekarství* 47, no. 1 (2001): 10–16.
17. SE O’Riordan et al., "Cystatin C Improves the Detection of Mild Renal Dysfunction in Older Patients," *Annals of Clinical Biochemistry* 40, no. 6 (2003): 648–655.
18. E Wasen et al., "Renal Impairment Associated with Diabetes in the Elderly," *Diabetes Care* 27, no. 11 (2004): 2648–2653.
19. BA Ozer et al., "Can Cystatin C Be a Better Marker for the Early Detection of Renal Damage in Primary Hypertensive Patients?" *Renal Failure* 27, no. 3 (2005): 247–253.
20. S Herget-Rosenthal et al., "Early Detection of Acute Renal Failure by Serum Cystatin C," *Kidney International* 66, no. 3 (2004): 1115–1122.
21. O Schuck et al., "Glomerular Filtration Rate Estimation in Patients with Advanced Chronic Renal Insufficiency Based on Serum Cystatin C Levels," *Nephron—Clinical Practice* 93, no. 4 (2003): c146–c151.
22. S Tian et al., "Cystatin C Measurement and Its Practical Use in Patients with Various Renal Diseases," *Clinical Nephrology* 48, no. 2 (1997): 104–108.

Copyright ©2006 IVD Technology



Copyright © 2005 Canon Communications, LLC [About Us](#) | [Advertising Information](#) | [Site Guide](#)