Critical Review Form
Clinical Prediction or Decision Rule

Kline JA, et al. Derivation and Validation of a Bayesian Network to Predict Pretest Probability of Venous Thromboembolism. Annals EM 2005; 45: 282-290.

Objective: To determine whether the Bayesian network can produce a pretest probability low enough to preclude D-dimer testing in ED patients. (p. 283)
Method: Bayesian networks apply intuitive reasoning methods to multiple variables simultaneously. Unlike a neural network, the structured output produced by a Bayesian network is interpretable, even if we cannot manage the complicated calculations without the computer. Perhaps the most important advantage of the Bayesian network is its ability to handle missing data by incorporating the most likely values of all the non-missing variables. In the current study, the authors derived a network using known values from 3,145 ED patients previously evaluated at 10 medical centers for venous thromboembolism. Using commercial software they then manipulated the probabilities of missing "nodes" (variables) and then repeated the analysis over 50 data sets (75 patients in each set) until those probabilities for each node which best predicted VTE were obtained. The model thus obtained was then tested on the prospective cohort.

	Guide	Comments
I.	***Is this a newly derived instrument* (Level IV)?**	Yes.
A.	Was validation restricted to the retrospective use of statistical techniques on the original database? (If so, this is a Level IV rule & is not ready for clinical application).	No, the Bayesian network was derived by secondary analysis of data collected prospectively on 3,145 ED patients at 10 hospitals in the US from 1996 to 2002" and "the best fit Bayesian network was then tested in 1,423 ED patients prospectively studied at Carolinas Medical Center and Brigham and Women's Hospital from January 1, 2001 to June 30, 2003." (pp 284-285). Thus, this study represents Level III evidence.
II.	**Has the instrument been validated? (Level II or III). If so, consider the following:**	
1a	Were all important predictors included in the derivation process?	The authors used a "custom data-mining tool that uses a genetic algorithm to search all possible Bayesian networks that can be developed from 25 clinical variables collected from the history and physical" (p. 283 and Figure 1 p. 289). **Five variables were excluded <u>because they did not contribute substantially to the determination of outcomes</u>**: systolic blood pressure, immobility, previous thromboembolism, hormone use, and pregnancy.

1b	Were all important predictors present in significant proportion of the study population?	Table 1 (p. 284) shows prevalence of variables among derivation and validation cohorts. Note is made of a few differences (cough, dyspnea, COPD, smoking status, and immobility), but no p-values are reported.
1c	Does the rule make clinical sense?	Although the technology is rather intimidating for technophobes (the second author is a computer science PhD!), the computerized CDR makes sense intellectually in that the authors study all known risk factors for VTE.
2	Did validation include prospective studies on several different populations from that used to derive it (II) or was it restricted to a single population (III)?	No, separate derivation and validation cohorts. See Ia answer.
3	*How well did the validation study meet the following criteria?*	
3a	Did the patients represent a wide spectrum of severity of disease?	All EM patients, presumably for a variety of complaints, so the results from this study may not be extrapolated to healthy outpatients.
3b	Was there a blinded assessment of the gold standard?	Not all patients had the "gold standard" in either derivation or validation cohort. However, "the composite criterion standard remains consistent with methodology used by other experts in the study of venous thromboembolism as reviewed by Kruip". (p. 284) On the validation set, criterion standard testing was only "performed if D-dimer testing results were abnormal". (p. 285). So Gold standard not always obtained and no mention of blinded reviewers of diagnostic studies.

3c	Was there an explicit and accurate interpretation of the predictor variables & the actual rule without knowledge of the outcome?	Presumably yes because the computer doing the math given only data input. In other words, although the computer is limited by the data it is given, it is not affected by the bias of humans (need to prove the CDR effective, publish results, etc.). So the computer was unlikely to recruit healthier patients or ignore unhelpful data as some humans (consciously or subconsciously) might do.

3d	Did the results of the assessment of the variables or of the rule influence the decision to perform the gold standard?	Unknown on derivation set. On validation set, D-dimer not one of 25 variables and that was only stated decision marker for further diagnostic testing.
4	How powerful is the rule (in terms of sensitivity & specificity; likelihood ratios; proportions with alternative outcomes; or relative risks or absolute outcome rates)?	The Bayesian network deemed half of the validation cohort as "low risk" which correctly identified 98.5% (700/711) of those with probability of DVT <2% (p. 286).
III.	**Has an impact analysis demonstrated change in clinical behavior or patient outcomes as a result of using the instrument? (Level I). If so, consider the following:**	
1	How well did the study guard against bias in terms of differences at the start (concealed randomization, adjustment in analysis) or as the study proceeded (blinding, co-intervention, loss to follow-up)?	No impact analysis was performed.
2	What was the impact on clinician behavior and patient-important outcomes?	No impact analysis was performed.

Limitations

1. Level III evidence.
2. Bayesian network not widely available or easy to understand for non-researchers.
3. Cannot extrapolate results in a population with a pre-test probability >11% (since not derived or validated in such a population).
4. Unable to assess reliability (Confidence Intervals) with present model.

Bottom Line

Further validation of a Bayesian network probably represents the future of clinical decision aids. Whereas typical CDR's provide a range of pre-test probability (low, moderate, high), Bayes theorem requires a pre-test point estimate in order to then utilize the test result and corresponding Likelihood Ratio to derive (mathematically or with Fagan's nomogram) a post-test probability. Most clinicians do not think in terms of a numeric pre-test point estimate and even those who do will note significant intra-rater (between self on similar presentations) and inter-rater (between rater 1 and rater 2) variability because they weigh variables differently based upon prior beliefs and experiences. The Bayesian network can categorically and objectively weigh risk factors and findings into a model that best fits your patients' characteristics, thus yielding a reproducible, reliable point estimate of disease probability (theoretically). When the network's probability is combined with your clinical intuition, both resource utilization and outcomes can be optimized. Although the long-term utility of Bayesian network technology remains uncertain, the implications of this technology deserve careful attention by EM physicians in coming years.