Critical Review Form

Clinical Decision Analysis

Bayes Pulmonary Embolism Assisted Diagnosis: A New Expert System for Clinical Use, *Emerg Med J* 2007; 24: 157-164

Objectives: **To test BAYPED (Bayes Pulmonary Embolism Assisted Diagnosis) can "help doctors in correctly classifying suspected thromboembolic events" (p 157) by describing diagnostic accuracy and correspondence between model estimated probability and the patient's true disease status.**

Methods: **Using the PISA-PED 1996 Italian database (Am J Respir Crit Care Med 1996; 154: 1387-1393) of 750 consecutive patients suspected by one of six experienced pulmonary physicians to have PE, the authors of the current study sought to develop a computer program utilizing any combination of 48 variables to recommend a next step. The cut-off for establishing PE as the diagnosis was a post-test probability > 95% while the cut-off to exclude PE as the diagnosis was <5% "provided that all the examinations who's costs divided by the probability of PE do not exceed conventional boundary of €3500". (p 158) Because the computer had the flexibility to calculate probabilities if information was missing (test requested not actually performed on the 1996 cohort), researchers felt the diagnostic system could adapt to the diagnostic resource needs of particular hospitals or wards.**

To measure how well predicted probabilities correlated with reality, the authors produced a plot of PE presence (DV) versus the ratio of probability ÷ (1-probability) as the Independent Variable via logistic regression analysis. If the accuracy of the model is perfect, the intercept would be zero and the slope one. An intercept other than zero would represent a systematic over- or under-representation of PE probability by the computer model.

The computer program was modified by accessing its calibration in 500 randomly selected cases from the PISA-PED database with refinements maintained only when they both improved diagnostic accuracy of the model AND "conformed to the medical literature". The accuracy was assessed with both including and excluding those who had confirmatory pulmonary angiography performed.

Guide		Comments
I.	**Are the results valid?**	
A.	**Were all important strategies and outcomes included?** *In other words, did the authors consider every potential course of action and possible outcome?*	No, because the computer program was tested on 1996 data. Duplex ultrasonography, D-dimer assays, and CT angiograms were often unavailable at that time. Furthermore, the models did not include Clinical Decision Rules, which might have further enhanced discriminatory abilities.
B.	**Was an explicit and sensible process used to identify, select, and combine the evidence into probabilities?** *In other words, the authors should perform as comprehensive a literature review as is required for a meta-analysis. In addition, probabilities must be assigned to each branch emanating from a chance node, and for each chance node, the sum of probabilities must add to 1.0.*	Although not well described in the methods and without any explicit description of where baseline probability or node probability values are, the extensive list of variables included in the model (Table 2, p 160) seem well-founded and best-evidence based.
C.	**Were the utilities obtained in an explicit and sensible way from credible sources?** *Utility represents the value to the patient of remaining expected life. A utility threshold of 0.92 means that your patient feels he would be willing to sacrifice 8% of his/her remaining life to avoid that limb of the decision tree (going on dialysis, taking Coumadin, etc.).* *In other words, were the quantitative measurements of the value to the decision maker of the various outcomes provided by someone who understands the outcomes and the condition being rated? Whatever the measurement method, the authors should report the source of the ratings. In a decision analysis built for an invidual patient, the most credible ratings are those measured directly from the patient.*	No patient-input or values were sought, thus no utility analysis was conducted.

D.	**Was the potential impact of any uncertainty in the evidence determined?** *Much of the uncertainty in clinical decision making arises from the lack of valid evidence in the literature. Even when present, published evidence is often imprecise with wide confidence intervals around estimates for important variables. Sensitivity analysis asks the question "Is the conclusion generated by the decision analysis affected by the uncertainties in our estimates of the likelihood of the outcomes?"* Satisfy yourself that all of the clinically important variables were included.	All clinically important variables were included, but the authors do not state any sensitivity analysis to account for the range of diagnostic accuracy for elements of history, physical exam, and labs/imaging.
II.	**What are the results?**	
A.	**In the baseline analysis, does one strategy result in a clinically important gain for patients? If not, is the result a "toss-up"?** *For a clinical decision analysis that compares two clinical strategies, there are three possible results: strategy 1 is better than strategy 2, strategy 2 is better than strategy 1, or both strategies are equally good or bad. A gain in life expectancy or quality-adjusted life expectancy of 2 or more months is considered an important gain.*	The best strategy (for the patient) is diagnosis of PE with minimal testing while avoiding testing those who do not have PE. In EM, we do struggle with a significant number of negative tests, so if the computer model can suggest preferential diagnostic testing non-invasively with equal or greater accuracy (to gestalt or CDR's) than a superior diagnostic tool will exist.
B.	**How strong is the evidence used in the analysis?** *Ideally, every probability estimate at every node in the tree is supported by precise estimates from primary and integrative studies of high methodological quality. The fewer the probabilities that can be precisely estimated from high quality primary studies, the weaker the overall inference one can make from the results.*	Although PE is a well-studied diagnostic conundrum, few precise estimates from primary high-quality studies exist. Regardless, the authors do not provide specific probabilities (or references for those probabilities) at each node.

C.	**Could the uncertainty in the evidence change the result?** *For any clinical variable the decision analyst can calculate the value or "threshold" above which the results favor one strategy and below which the results favor another strategy. If the result of the analysis would change by choosing different values for one of the variables, the result is said to be "sensitive" to that variable.*	Unknown since the authors do not conduct a sensitivity analysis.

<table>
<tr><td>III.</td><td>Will the results help me in caring for my patients?</td><td></td></tr>
<tr><td>A.</td><td>Do the probability estimates fit my patients' clinical features?

If the analysis was intended for patients different from yours, review the results of the sensitivity analyses. If the clinical characteristics of the intended patients are different from yours, you should discard the results. If a clinical decision analysis shows that the preferred strategy is sensitive to a given variable, you will need to gauge where your patient fits on the scale of that variable.</td><td>Although the PESA-PED data set is Italian patients, no good reason exists to suggest different patient characteristic or diagnostic test performance between Italian and St. Louis PE patients. Furthermore, there is no diagnostic literature suggesting differential diagnostic features of PE based upon ethnicity.

The current study provides the following probability results:

• PE prevalence of 40.1% in the sample of 500 subjects upon which BAYPED was calibrated.

Pre-refinement BAYPED performance:
PE
<table><tr><th>BAYPED</th><th>+</th><th>-</th></tr><tr><td>+</td><td>185</td><td>40</td></tr><tr><td>-</td><td>17</td><td>258</td></tr></table>Sensitivity = 92%
Specificity = 87%
LR+ = 6.8 (5.1-9.1)
LR- = 0.10 (0.06 – 0.15)
Accuracy = 88.6%

Refined BAYPED model:
PE
<table><tr><th>BAYPED</th><th>+</th><th>-</th></tr><tr><td>+</td><td>196</td><td>8</td></tr><tr><td>-</td><td>6</td><td>290</td></tr></table>Sensitivity = 97%
Specificity = 97%
LR+ = 36 (18-72)
LR- = 0.03 (0.01-0.07)
Accuracy = 96%</td></tr>
</table>

		• The refined model had an intercept of -0.05 and slope of 0.62 demonstrating no significant over/under estimation of PE risk, but imperfect correlation with actual PE prevalence.
B.	**Do the utilities reflect how my patients would value the outcomes of the decision?** *You must consider whether your patient's values are similar to those used in the decision analysis. If you were to ask your patient to rate the outcome states using the rating instrument in the article, you would know exactly what utility values to use.*	Presumably patients would rate accurate, efficient diagnosis (or exclusions) of PE very important, but future studies of this computer aid should confirm this assumption.
C.	**Can I use the results in the management of patients in my practice?**	This is the million dollar (€ 735,781) question. Who will input this data? Can it be made compatible with electronic medical records like HMED? How expensive will it be? Will clinicians accept such a diagnostic aid? Will patients?

Limitations

1) Complicated model with insufficient detail about model probabilities (values, references).

2) No utility assessment of patient values (or clinician acceptability).

3) Dated validation sample lacks use of Doppler ultrasound, D-dimer, PE protocol CT, or Clinical Decision Rules perhaps over-representing the recommendation of pulmonary angiogram in one-third of subjects.

4) Authors fail to elaborate on specific changes to the computerized algorithm through their refinements.

5) No discussion of expense of this system or integration feasibility with existing electronic medical record systems.

Bottom Line

Computerized Bayesian decision models offer a promising, if somewhat intellectually challenging, diagnostic assistance device for complicated diagnoses like PE. Because presentations are myriad, CDR components subjective and sporadically unavailable (ABG for Geneva CDR), and individual diagnostic elements inter-related (surgery and D-dimer utility), computerized models offer a flexibility advantage while systematically, non-subjectively, quantitatively assigning risk. Much research is required in the future (reproducibility in different health care settings, ease of data acquisition, clinician/patient acceptability), but given limitations of current diagnostic strategies and the dawn of electronic medical records, computerized Bayesian decision models likely represent 21st Century medicine Standard of Care.