Prediction Rules Must Be Developed According to Methodological Guidelines

TO THE EDITOR: We read with interest the article by Chan and colleagues (1). They described development of a prediction rule that could be used for pretest probability assessment in pregnant women with suspected deep venous thrombosis (DVT). Prediction rules are valuable tools in daily clinical practice because they provide absolute risks for individual patients, and these risks can be used to determine treatment choices or further diagnostic work-up. However, before a prediction rule can be used safely in daily clinical practice, the methodological soundness of the development steps needs to be assured. Chan and colleagues made some methodological choices that we disagree with.

First, the authors included predictors in the prediction rule if they increased the c-statistic by at least 0.03. However, this approach has been severely criticized, because it is less sensitive than measures based on the likelihood ratio model chi-square or other global measures of model fit (2–4). The authors should have used the likelihood ratio model chi-square to include predictors in the model, if predictor selection is to be used at all (not advised).

Second, the authors did a univariate analysis and entered predictors that were significantly associated with presence of DVT at a P value of 0.05 or less into a multivariate model. Selecting predictors at such a low P value has been strongly advised against, because it leads to poor model performance when evaluated on new patients (external validation), especially in small data sets (5, 6). This is particularly important in this data set because the authors explored too many predictors, considering the low number of events: 11 predictors on 17 events. The authors acknowledged that they needed at least 5 to 10 events per predictor when they developed the rule. They state that, using an initial model with 6 predictors, they needed between 30 and 60 events. However, the authors did not account for the univariate step when making this calculation. Taking this analysis into account, 55 to 110 events would have been needed for the initial 11 predictors that were analyzed.

Third, the authors did not use any imputation method to handle the missing values in their data set. This means that they performed a complete case analysis and analyzed only the data of the patients with complete records. One of the predictors in the model (difference in calf circumference) was missing in 46 of the 194 patients, which means the information of only 75% of the patients was analyzed. Besides a loss of power, it is widely acknowledged that ignoring the missing values in a data set by conducting a complete case analysis may bias study results (7, 8).

Prediction rules are valuable tools for daily clinical practice. Not surprisingly, the number of articles on prediction rules in the medical literature has increased enormously (it more than doubled between 1995 and 2005) (9). Most of the articles concern development of prediction rules. Considering this rapid increase, we believe these prediction rules should be developed according to methodological guidelines. We therefore recommend that researchers developing prediction rules use these guidelines (4–6, 10–13). A selection of the recommended references can also be found in the Information for Authors section of the Annals Web site (14).

Krisel J.M. Jansen, PhD
Karel G.M. Moons, PhD
University of Utrecht
3508 Utrecht, the Netherlands

Frank E. Harrell Jr., PhD
Vanderbilt University Medical School
Nashville, TN 37232

Potential Conflicts of Interest: None disclosed.

References

IN RESPONSE: We appreciate Dr. Jansen and colleagues’ comments. They raise some important and valid concerns, which we address below.

As Dr. Jansen and colleagues point out, prediction rules are important to guide clinical practice in many instances, because they provide absolute risks for individual patients that can be used to determine treatment choices or further diagnostic work-up. We agree that before any rule can be adopted, it must be properly validated. However, before a prediction rule can be used safely in daily clinical practice, the methodological soundness of the development steps needs to be assured. Chan and colleagues made some methodological choices that we disagree with.

First, the authors included predictors in the prediction rule if they increased the c-statistic by at least 0.03. However, this approach has been severely criticized, because it is less sensitive than measures based on the likelihood ratio model chi-square or other global measures of model fit (2–4). The authors should have used the likelihood ratio model chi-square to include predictors in the model, if predictor selection is to be used at all (not advised).

Second, the authors did a univariate analysis and entered predictors that were significantly associated with presence of DVT at a P value of 0.05 or less into a multivariate model. Selecting predictors at such a low P value has been strongly advised against, because it leads to poor model performance when evaluated on new patients (external validation), especially in small data sets (5, 6). This is particularly important in this data set because the authors explored too many predictors, considering the low number of events: 11 predictors on 17 events. The authors acknowledged that they needed at least 5 to 10 events per predictor when they developed the rule. They state that, using an initial model with 6 predictors, they needed between 30 and 60 events. However, the authors did not account for the univariate step when making this calculation. Taking this analysis into account, 55 to 110 events would have been needed for the initial 11 predictors that were analyzed.

Third, the authors did not use any imputation method to handle the missing values in their data set. This means that they performed a complete case analysis and analyzed only the data of the patients with complete records. One of the predictors in the model (difference in calf circumference) was missing in 46 of the 194 patients, which means the information of only 75% of the patients was analyzed. Besides a loss of power, it is widely acknowledged that ignoring the missing values in a data set by conducting a complete case analysis may bias study results (7, 8).

Prediction rules are valuable tools for daily clinical practice. Not surprisingly, the number of articles on prediction rules in the medical literature has increased enormously (it more than doubled between 1995 and 2005) (9). Most of the articles concern development of prediction rules. Considering this rapid increase, we believe these prediction rules should be developed according to methodological guidelines. We therefore recommend that researchers developing prediction rules use these guidelines (4–6, 10–13). A selection of the recommended references can also be found in the Information for Authors section of the Annals Web site (14).

Letter from Frank E. Harrell Jr., PhD
under the receiver-operating characteristic curve), it is unlikely that any other significant predictors were missed by using this method.

We also agree with Dr. Janssen and colleagues’ second point. The low event rate in our study reflects the reality seen in the few diagnostic studies of DVT in pregnant women. The prevalence of DVT in this particular cohort of patients is low (<10%). Despite our best efforts over 7 years, we were limited to 17 events (out of 194 patients). The rule that one should have 5 to 10 events per variable in a multivariable logistic model is based on the fact that having fewer events leads to unstable variable estimates. Thus, it would have been unreasonable to attempt to fit a single model with 11 independent variables to our data, but this does not imply that there is a problem with fitting 11 single-variable models.

The last comment by Dr. Janssen and colleagues is also valid. Our 2 objectives were to determine how clinicians find DVT in pregnant women by subjective means and whether any “objective” predictors exist that can help clinicians to do this. We believe that we have achieved both objectives. We share all the concerns raised by Dr. Janssen and colleagues, including those regarding the development of prediction rules. We repeatedly emphasized throughout our article that this rule should not be applied in daily practice until it has been properly validated.

In pregnant women, symptoms mimicking DVT are common (for example, leg swelling and pain). At the very least, our study will increase awareness that when certain symptoms (for example, left leg presentation and asymmetry) are present, one should be more vigilant for the presence of DVT and arrange for appropriate testing.

Whee Shian Chan, MD, MSc
Women’s College Hospital
Toronto, Ontario M5G 1B2, Canada

Tim Ramsay, MSc, PhD
Ottawa Health Research Institute
Ottawa, Ontario K1Y 4E9, Canada

Jeffrey S. Ginsberg, MD
McMaster University Medical Centre
Hamilton, Ontario L8N 3Z5, Canada

Potential Conflicts of Interest: None disclosed.

The Net Clinical Benefit of Warfarin Anticoagulation in Atrial Fibrillation

TO THE EDITOR: With regard to the article by Singer and colleagues (1), calculating net clinical benefit of warfarin therapy in atrial fibrillation is an important concept (2). However, 3 other issues should be considered in the future.

First, warfarin therapy affects other major events and may substantially reduce myocardial infarction and death (3).

Second, because control of the international normalized ratio (INR) varies in different settings and because a 10% improvement in the amount of time the INR is in the therapeutic range (TTR) has been associated with a greater than 10% reduction in event rates (3, 4), the relationship between INR control and event rates must be more thoroughly defined before it can be determined how INR control may affect the net clinical benefit in one’s own setting. Veeger and colleagues (5) found that the bottom quartile of approximately 4000 patients was in range only 10% to 20% of the time and accounted for more than half of the major events. Jones and colleagues (4) found that the bottom quartile of 2223 patients was in range only about 28% of the time. Furthermore, because event rates increase exponentially as the INR moves further out of range, being slightly out of the target range may have little effect, whereas being at the extremes (INR, <1.5 or >5.0) may carry a very high risk. Therefore, one must know what the event rates were when the INR was in the target range TTR = 0.3 INR units, below an INR of 1.5, and above an INR of 5.0 to estimate the net clinical benefit in one’s own setting.

Third, one must consider how evolving methods to improve INR control may alter the net clinical benefit. As noted by Hart and Halperin (2), the TTR of 65% reported by Singer and colleagues is often considered “high.” By combining INR self-testing and computer management, however, Harper and Pollock (6) reported an 80% TTR with no INRs greater than 5. An interim analysis of our similar study (7) found a TTR of 78.9%, which increased to 94% when the range was expanded slightly by ±0.3 INR units. Approximate TTR values were 90% for the top quartile and 60% for the lowest quartile, and the percentage of time that the INR was greater than 5 or less than 1.5 was only 0.27%. Such improved INR control is estimated to yield a 30% to 50% reduction in both thromboembolic and major bleeding events compared with “typical” management—changes that should have a substantial effect on the net clinical benefit of warfarin.

Henry Bussey, PharmD
The University of Texas at Austin
Austin, TX 78712

Potential Conflicts of Interest: Consultancies: Genesis Advanced Technologies. Stock ownership or options (other than mutual funds): Inverness Medical. Patents received/pending: 3 vitamin K antagonist products.

References
7. Bussey HI, Walker MB, Bussey-Smith KL, Frei CR. Interim analysis of triple...