Meta-analysis of clinical prediction models

Thomas P.A. Debray, PhD
Introduction

"We demand rigidly defined areas of doubt and uncertainty!"
Douglas Adams, *The Hitchhiker’s Guide to the Galaxy*
Clinical Prediction Models

• Estimation of absolute risk using multiple predictors
  – Demographic characteristics
  – Clinical history and physical examination
  – Medical imaging, elektrofysiology, pathology
  – Biomarker tests

• Diagnostic prediction models
  – Predict presence of a certain disease or condition

• Prognostic prediction models
  – Predict future occurrence of a certain outcome

• Guide healthcare professionals and individuals
Clinical Prediction Models

• Developed from subject-level data
  – Cross-sectional studies (diagnostic models)
  – Prospective cohort study (prognostic models)
  – (Other designs)

• Statistical data analyses
  – Data cleaning
  – Predictor selection
  – Missing data

• Presentation
  – Equations to calculate outcome risk
  – Score charts
Deep Venous Thrombosis (DVT)

- Blood clot that forms in a vein in the body (lower leg/thigh)
- If blood clot breaks off -> blood stream -> lungs -> blockage
- Pulmonary embolism, preventing oxygenation of blood
- Potentially causing death
Diagnosis of DVT

- Limited value of signs and symptoms in primary care
- Most patients suspected of DVT referred to secondary care
- Reference standard: ultrasonography (CUS)
- Burden on patients and health care budgets

Need for developing **multivariable prediction models**
- Predict presence of DVT in suspected patients
  - Patient history and physical examination
  - Biomarker test results: D-dimer test
- Primary care versus secondary care
# Diagnosis of DVT

## Wells Score (DVT)

<table>
<thead>
<tr>
<th>Factor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active cancer (treatment ongoing or within previous 6 months, or palliative treatment)</td>
<td>1</td>
</tr>
<tr>
<td>Paralysis, paresis, or recent plaster immobilization of the lower extremities</td>
<td>1</td>
</tr>
<tr>
<td>Recently bedridden for 3 days or more, or major surgery within the previous 12 weeks requiring general or regional anesthesia</td>
<td>1</td>
</tr>
<tr>
<td>Localized tenderness along the distribution of the deep venous system</td>
<td>1</td>
</tr>
<tr>
<td>Entire leg swollen</td>
<td>1</td>
</tr>
<tr>
<td>Calf swelling &gt; 3 cm compared to asymptomatic leg (measuring 10 cm below tibial tuberosity)</td>
<td>1</td>
</tr>
<tr>
<td>Pitting edema confined to the symptomatic leg</td>
<td>1</td>
</tr>
<tr>
<td>Nonvaricose collateral superficial veins</td>
<td>1</td>
</tr>
<tr>
<td>Previously documented DVT</td>
<td>1</td>
</tr>
<tr>
<td>Alternative diagnosis at least as likely as DVT</td>
<td>-2</td>
</tr>
</tbody>
</table>

- **≤ 0**: LOW pretest probability
- **1 or 2**: MODERATE pretest probability
- **≥ 3**: HIGH pretest probability


*In patients with symptoms in both legs, the more symptomatic leg is used.*
## Diagnosis of DVT

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Hamilton</th>
<th>Modified Wells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plaster immobilization of lower limb</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Active malignancy (within 6 months or current)</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Strong clinical suspicion of deep venous thrombosis by the emergency physicians without other diagnostic possibilities</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>Bed rest (&gt;3 days) or recent surgery (within 4 weeks)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Male sex</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Calf circumference &gt;3 cm on affected side (measured 10 cm below tibial tuberosity)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Erythema</td>
<td>1</td>
<td>–</td>
</tr>
<tr>
<td>Localized tenderness along the distribution of the deep venous system</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Entire leg swollen</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Pitting edema confined to the symptomatic leg</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Collateral superficial veins (nonvaricose)</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Previously documented deep vein thrombosis</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Alternative diagnosis at least as likely as deep vein thrombosis</td>
<td>–</td>
<td>−2</td>
</tr>
<tr>
<td>Unlikely versus likely cutoff score</td>
<td>2 or less</td>
<td>1 or less</td>
</tr>
</tbody>
</table>
# Diagnosis of DVT

<table>
<thead>
<tr>
<th>Diagnostic variables</th>
<th>Odds ratio</th>
<th>Regression coefficient*</th>
<th>p-value</th>
<th>Points for the rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male gender</td>
<td>1.80 (1.36 – 2.16)</td>
<td>0.59</td>
<td>&lt;0.001</td>
<td>1</td>
</tr>
<tr>
<td>Oral contraceptive use</td>
<td>2.12 (1.32 – 3.35)</td>
<td>0.75</td>
<td>0.002</td>
<td>1</td>
</tr>
<tr>
<td>Presence of malignancy</td>
<td>1.52 (1.05 – 2.44)</td>
<td>0.42</td>
<td>0.082</td>
<td>1</td>
</tr>
<tr>
<td>Recent surgery</td>
<td>1.46 (1.02 – 2.09)</td>
<td>0.38</td>
<td>0.044</td>
<td>1</td>
</tr>
<tr>
<td>Absence of leg trauma</td>
<td>1.82 (1.25 – 2.66)</td>
<td>0.60</td>
<td>0.002</td>
<td>1</td>
</tr>
<tr>
<td>Vein distension</td>
<td>1.62 (1.19 – 2.20)</td>
<td>0.48</td>
<td>0.002</td>
<td>1</td>
</tr>
<tr>
<td>Calf difference ≥ 3 cm</td>
<td>3.10 (2.36 – 4.06)</td>
<td>1.13</td>
<td>&lt;0.001</td>
<td>2</td>
</tr>
<tr>
<td>D-dimer abnormal</td>
<td>20.3 (8.25 – 49.9)</td>
<td>3.01</td>
<td>&lt;0.001</td>
<td>6</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DVT = deep vein thrombosis; *=natural logarithm of the odds ratio; D-dimer abnormal for VlDAS ≥ 500 ng/ml and Tinaquant ≥ 400 ng/ml. Probability of DVT as estimated by the final model:

\[
\text{P(DVT)} = \frac{1}{1 + \exp(-5.47 + 0.59 \times \text{male gender} + 0.75 \times \text{OC use} + 0.42 \times \text{presence of malignancy} + 0.38 \times \text{recent surgery} + 0.60 \times \text{absence of leg trauma} + 0.48 \times \text{vein distension} + 1.13 \times \text{calf difference ≥ 3 cm} + 3.01 \times \text{abnormal D-dimer})}
\]
## Tableau II

Analyse multivariée : modèle de régression logistique final prédisant la présence d’une thrombose veineuse profonde

<table>
<thead>
<tr>
<th>Variable</th>
<th>p</th>
<th>odds ratio</th>
<th>coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immobilisation médicale dans le mois précédent (aliment &gt; 48 h ou paralysie)</td>
<td>0,07</td>
<td>1,9 (1,0–3,7)</td>
<td>0,62</td>
</tr>
<tr>
<td>Contraception oestroprogestative</td>
<td>0,02</td>
<td>4,0 (1,2–12,9)</td>
<td>1,38</td>
</tr>
<tr>
<td>Antécédent personnel de MVTE</td>
<td>0,02</td>
<td>2,1 (1,1–4,0)</td>
<td>0,74</td>
</tr>
<tr>
<td>Cancer évolutif</td>
<td>&lt;0,01</td>
<td>7,3 (2,4–22,1)</td>
<td>1,99</td>
</tr>
<tr>
<td>Diminution du ballant du mollet</td>
<td>0,01</td>
<td>2,3 (1,3–4,1)</td>
<td>0,83</td>
</tr>
<tr>
<td>Diagnostic alternatif au moins aussi probable</td>
<td>&lt;0,01</td>
<td>0,1 (0,1–0,3)</td>
<td>−2,08</td>
</tr>
</tbody>
</table>
External validation

- Which model should we use?
- What performance can we expect?
- Does the model require improvements/changes?
- Or, should we rather develop a model from scratch?

External validation is needed!
- Identify and evaluate existing models
- Assess performance in a new sample
- Compare predicted probabilities to observed outcomes
- Distinguish between discrimination and calibration
External validation

Validation sample for DVT models

- Prospective management study
- 300 primary care practices in 3 regions of the Netherlands (Amsterdam, Maastricht, and Utrecht)
- Outcome: incidence of symptomatic venous thromboembolism during 3-month follow-up
- 1028 patients with clinically suspected DVT
- 131/1028 patients eventually diagnosed with DVT

Question: Can the previously identified models predict which subjects have DVT?
External validation (Gagne)

Discrimination

AUC = 0.81

Calibration

Actual probability vs. Predicted probability
External validation

• Discrimination secondary care models
  – 0.66 (Hamilton)
  – 0.76 (Wells)
  – 0.77 (modified Wells)

• Discrimination primary care models
  – 0.81 (Gagne)
  – 0.82 (Oudega)

**Remark:** Secondary-care models may not adequately rule out DVT in primary-care settings!
The Wells Rule Does Not Adequately Rule Out Deep Venous Thrombosis in Primary Care Patients

Ruud Oudega, MD; Arno W. Hoes, MD, PhD; and Karel G.M. Moons, PhD

[+] Article and Author Information

Companion Article(s):
Meta-Analysis: The Value of Clinical Assessment in the Diagnosis of Deep Venous Thrombosis
Use of a Clinical Prediction Score in Patients with Suspected Deep Venous Thrombosis: Two Steps Forward, One Step Back?

Model updating

Adjust promising models to the validation sample

- Adjust intercept
correct for different outcome prevalence
- Adjust intercept and common slope
correct for different outcome prevalence and predictor
effects that are over-optimistic
- More advanced updating procedures
  - Adjust a particular regression coefficient
  - Re-estimate all regression coefficients
  - Add completely new predictors

**Remark:** updating procedures reduce insight into model validity as new parameters are being estimated
Model updating (Gagne)

Update of intercept and common slope
Caveats of prediction modeling research

- Most models are never validated
- Model redevelopment versus model updating
- Risk of overfitting
- Prior knowledge not optimally used
- Incompatibility and confusion

The user must typically choose between a cacophony of existing models for which performance may be obscure
Meta-analysis of prediction models

• Meta-analysis (therapeutic research)
  – Synthesize evidence from multiple trials
  – Obtain a summary estimate of treatment effect
  – Facilitate detailed analyses of effect modification

• Meta-analysis (prognostic research)
  – Synthesize evidence on prognostic factors
  – Aggregate literature models into a meta-model that is optimized for validation sample
  – Improve generalizability of meta-model across different patient populations

• How to combine models with similar predictors?
• How to combine models with different predictors?
Meta-analysis of prediction models
Meta-analysis of prediction models

Aggregation of prediction models with similar predictors

• Identify common predictors
  – restore missing coefficients and standard errors where necessary (imputation)

• Pooling of predictor effects
  – calculate weighted average of regression coefficients
  – account for differences in precision
  – account for heterogeneity across studies

• Meta-model for average or specific study population
  – Relevance of literature versus validation sample
  – Adjust intercept term to local circumstances
Meta-analysis of prediction models

Aggregation of prediction models with similar predictors

- Univariate meta-analysis
  - pool predictor effects separately
- Multivariate meta-analysis
  - simultaneous pooling of all predictor effects
- Multivariate meta-analysis + Bayesian inference
  - pooled predictor effects from the literature are used as prior information for the predictor effects in the validation sample
Meta-analysis of prediction models

Aggregation of prediction models with similar predictors

• Diagnosis of DVT: focus on 4 common core predictors (+ intercept term)
Meta-analysis of prediction models

Multivariate meta-analysis

Bayesian inference

AUC = 0.8

Predicted probability

Actual probability
Meta-analysis of prediction models

Aggregation of prediction models with similar predictors

• (Simplified) meta-model
  – fewer predictors
  – adjusted for validation sample (baseline risk)
  – similar performance as best literature model

Implementation difficult when literature models differ much in terms of included predictors
Meta-analysis of prediction models
Meta-analysis of prediction models

Aggregation of models with different predictors

• Model averaging
  1. Update literature models
  2. Calculate predictions for each subject, for each model
  3. Evaluate performance literature models
  4. Calculate weights based on model fit and updating complexity (BIC)
  5. Obtain (weighted) average predictions
  6. Calculate summary model
Meta-analysis of prediction models

Aggregation of models with different predictors

- DVT case study
  - update intercept and common slope of all models
  - Weights: 0.998 (Oudega), 0.002 (Gagne), 0 (other models)
Meta-analysis of prediction models

Aggregation of models with different predictors

• Stacked regressions
  – Weight predictions from literature models
  – Discard models with little (added) value
  – Update common intercept and overall slope
  – No distinct steps, one straightforward estimation procedure
  – Borrows less information from validation sample (as compared to model averaging)
Meta-analysis of prediction models

Aggregation of models with different predictors

- DVT case study
  - Weights: 1.01 (α), 0.537 (Oudega), 0.497 (Gagne), 0 (other models)
Meta-analysis of prediction models

Model Averaging

AUC = 0.82

Stacked Regressions

AUC = 0.85
Meta-analysis of prediction models

External validation of meta-models

- **Primary Care (N=791)**
  - Best literature model: AUC = 0.77, slope = 1.13
  - Model Averaging: AUC = 0.77, slope = 1.13
  - Stacked Regressions: AUC = 0.74, slope = 0.82

- **Secondary Care (N=1756)**
  - Best literature model: AUC = 0.84, slope = 1.29
  - Model Averaging: AUC = 0.86, slope = 1.29
  - Stacked Regressions: AUC = 0.88, slope = 1.33

Meta-model outperforms existing models for primary and secondary care settings!
Meta-analysis of prediction models

Simulation studies

• Model re-development only useful when
  – Large (validation) sample available
  – Literature models too heterogeneous with target population (i.e. differences beyond intercept and common slope)

• For small (validation) samples:
  – Model redevelopment techniques (e.g. backward selection or PMLE) outperformed by meta-analysis
  – Model updating techniques outperformed by meta-analysis
Discussion

• Novel paradigm for model development & validation
• Model aggregation versus selective updating
• Better use of prior knowledge, but only if relevant for target population

• Future research
  – Quality appraisal of literature models
  – Alternative weighting schemes
  – Mixed sources of literature evidence
  – Variable selection