Clinical Prediction Modeling and External Knowledge

Thomas Debray

Julius Center for Health Sciences and Primary Care
University Medical Center Utrecht
The Netherlands

January 11, 2012

University Medical Center Utrecht

Julius Center for Health Sciences and Primary Care

Clinical Prediction Modeling

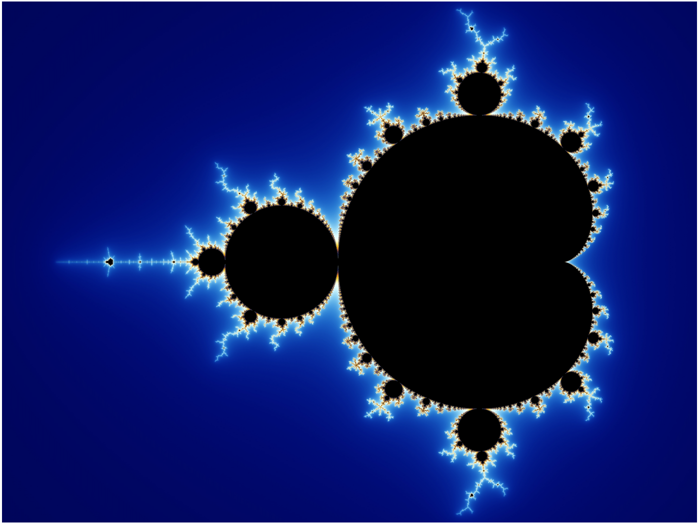
Clinical prediction models aim to provide a probability of disease/outcome presence (diagnosis) or occurrence (prognosis) in an individual.

Example: The Wells Rule

Clinical feature	Score
Active cancer (treatment ongoing or within previous 6 months or palliative)	1
Paralysis, paresis, or recent plaster immobilisation of the lower extremities	1
Recently bedridden for more than 3 days or major surgery, within 4 weeks	1
Localised tenderness along the distribution of the deep venous system	1
Entire leg swollen	1
Calf swelling by more than 3 cm when compared with the asymptomatic leg (measured 10 cm below tibial tuberosity)	1
Pitting oedema (greater in the symptomatic leg)	1
Collateral superficial veins (non-varicose)	1
Alternative diagnosis as likely or greater than that of deep-vein thrombosis	−2

In patients with symptoms in both legs, the more symptomatic leg is used.

Table 1: **Clinical model for predicting pretest probability for deep-vein thrombosis**



Evaluating generalizability

Important components

- Reproducibility
- Transportability

Validation studies

- Internal validation
- Temporal validation
- External validation

Search for **different but related** populations!

Example: External validation of the Wells Rule

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

Background: Using data from secondary care outpatients, Wells and colleagues developed a diagnostic rule to estimate the probability of the presence of deep venous thrombosis (DVT). The accuracy of the Wells rule has not been properly validated for use in primary care patients in whom DVT is suspected.

Objective: To validate the diagnostic accuracy of the Wells rule, with and without D-dimer testing, in a primary care setting.

Design: Cross-sectional study with prospective data collection from 1 January 2002 to 1 March 2003.

Setting: 110 primary care practices in a circumscribed geographic region in The Netherlands.

Participants: 1295 consecutive patients who consulted their primary care physician about symptoms suggestive of DVT.

Measurements: All patients underwent history-taking and physical examination to calculate the Wells rule score, and D-dimer testing. Repeated leg ultrasonography was the reference standard to determine the true presence or absence of DVT.

Results: In the primary care setting, 12.0% of patients in the low-risk group had DVT; the original study by Wells and colleagues reported a rate of 3% among such patients. When combined with negative results on a D-dimer test, the Wells rule yielded a prevalence of DVT of 2.9% in the lowest-risk group, whereas the prevalence was 0.9% in the original study.

Limitations: Patients with previous DVT were included, and the diagnostic reference standard was different from that used in Wells and colleagues' original study.

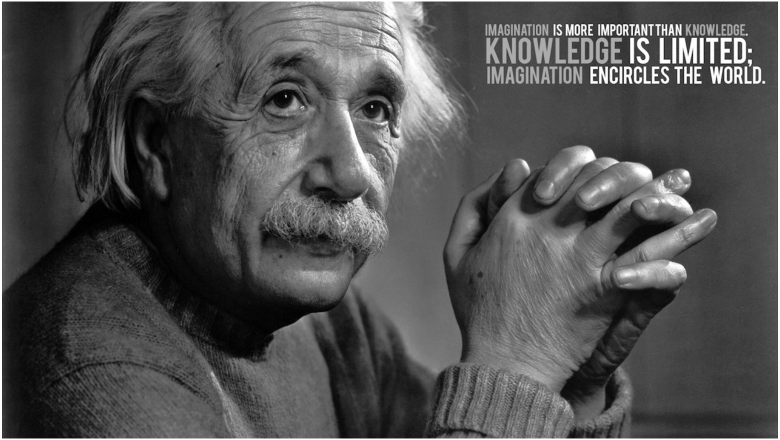
Conclusion: The Wells rule, alone or in combination with D-dimer testing, does not guarantee accurate estimation of risk in primary care patients in whom DVT is suspected.

Ann Intern Med. 2005;143:100-107. www.annals.org
For author affiliations, see end of text.

Improving generalizability

- Increase Sample Size
  - Individual Participant Data
  - Individual Study Centers
- Amplify Sample Spectrum
  - Domain
  - Heterogeneity
- Apply Robust Estimation
  - Penalization & Shrinkage
  - Model Updating
  - Including **External Knowledge**

"Can you increase its range?"

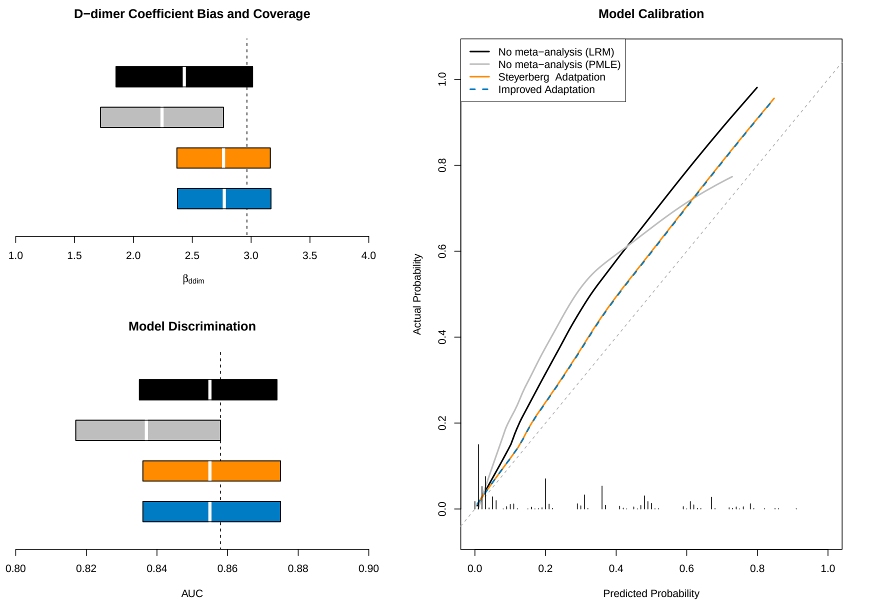

IMAGINATION IS MORE IMPORTANT THAN KNOWLEDGE.
KNOWLEDGE IS LIMITED;
IMAGINATION ENCIRCLES THE WORLD.

Overview of External Knowledge

- Available from literature (**binary outcomes**):
  - Univariable logistic regression coefficients (or unadjusted odds ratios)
  - Multivariable logistic regression coefficients (or adjusted odds ratios)
  - Complete logistic regression models (or score charts)
  - Regression trees, neural networks, ...
- Challenges
  - Heterogeneity !!!
  - Variety in modeling algorithms
  - Variety in considered parameters

External Knowledge: univariable regression coefficients

- **Adaptation Method**: update multivariable coefficient
  - Calculates change from uni- to multivariable coefficient
  - Applies change to (summarized) literature coefficients
  - Accounts for correlations *
  - Penalizes the adaptation *
- **Practical Example**: diagnosing DVT
  - IPD: Multivariable dataset ($n = 1295$)
  - LIT: 7 unadjusted odds ratios (biomarker D-dimer)
  - Update D-dimer coefficient in multivariable prediction model
  - External validation of updated prediction model ($n = 1756$)

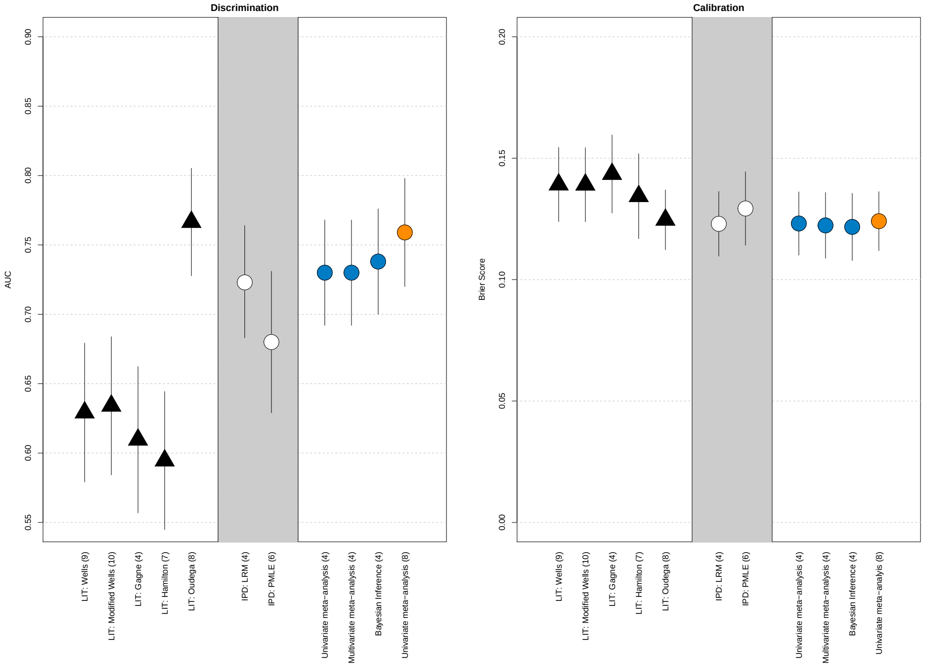

D−dimer Coefficient Bias and Coverage
1.0
1.5
2.0
2.5
3.0
3.5
4.0
β_{ddim}
Model Discrimination
0.80
0.82
0.84
0.86
0.88
0.90
AUC
Model Calibration
No meta−analysis (LRM)
No meta−analysis (PMLE)
Steyerberg Adatpation
Improved Adaptation
Actual Probability
0.0
0.2
0.4
0.6
0.8
1.0
Predicted Probability

External Knowledge: logistic regression models

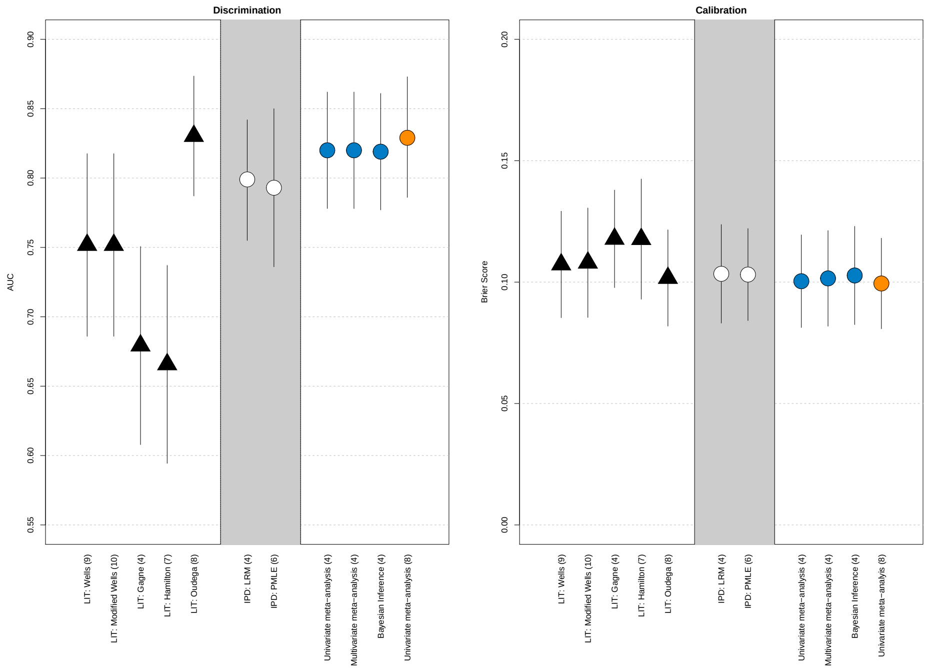
- Assumptions
  - Logistic regression models
  - Similar set of predictors
- Relaxing **constraints**
  - Focus on core set of predictors
  - Imputation of unknown regression coefficients
  - Recalculation of regression coefficients (score charts)
- Combining regression coefficients (LIT+IPD)
  - Univariate meta-analysis
  - Multivariate meta-analysis
  - Bayesian Inference
- Intercept update using IPD

External Knowledge: logistic regression models

- **Practical Example**: diagnosing DVT
  - IPD: Multivariable dataset ($n = 1028$)
  - LIT: 5 previously published prediction models
  - VAL 1: Multivariable dataset ($n = 791$)
  - VAL 2: Multivariable dataset ($n = 436$)
- **Constraints** literature models
  - Unknown standard errors
  - Unknown regression coefficients
  - Heterogeneous sets of predictors
  - Heterogeneous populations



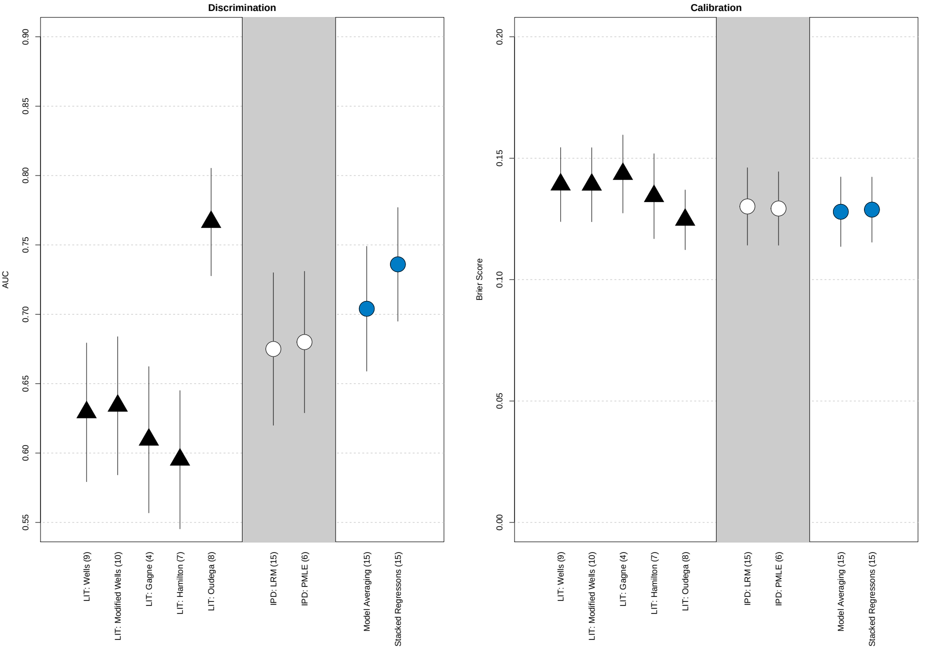
Discrimination
Calibration
AUC
Brier Score
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.20
0.15
0.10
0.05
0.00
LIT: Wells (9)
LIT: Modified Wells (10)
LIT: Gagne (4)
LIT: Hamilton (7)
LIT: Oudega (8)
IPD: LRM (4)
IPD: PMLE (6)
Univariate meta−analysis (4)
Multivariate meta−analysis (4)
Bayesian Inference (4)
Univariate meta−analysis (8)
LIT: Wells (9)
LIT: Modified Wells (10)
LIT: Gagne (4)
LIT: Hamilton (7)
LIT: Oudega (8)
IPD: LRM (4)
IPD: PMLE (6)
Univariate meta−analysis (4)
Multivariate meta−analysis (4)
Bayesian Inference (4)
Univariate meta−analyis (8)



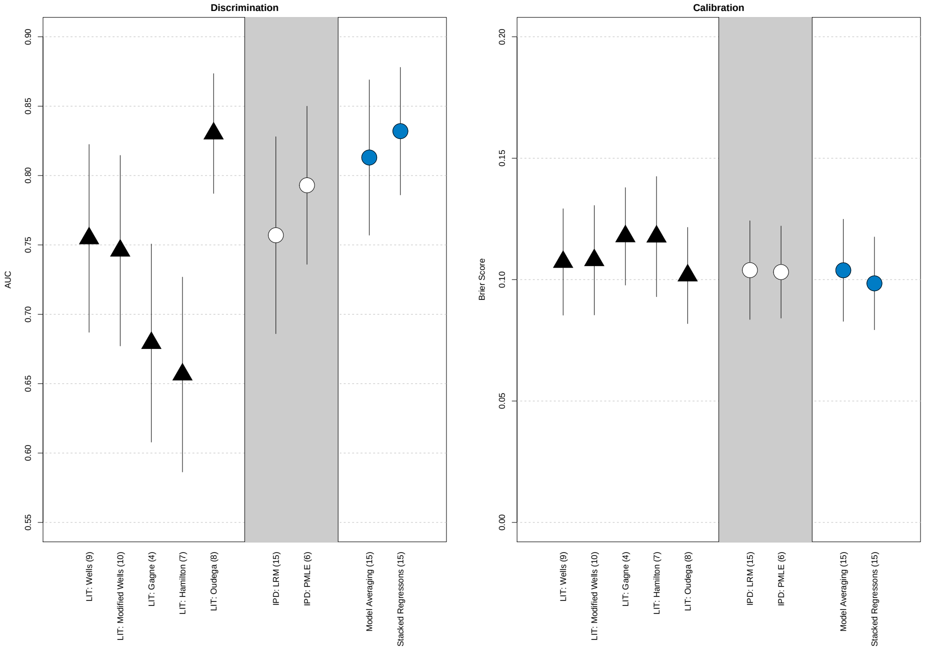
Discrimination
AUC
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
LIT: Wells (9)
LIT: Modified Wells (10)
LIT: Gagne (4)
LIT: Hamilton (7)
LIT: Oudega (8)
IPD: LRM (4)
IPD: PMLE (6)
Univariate meta−analysis (4)
Multivariate meta−analysis (4)
Bayesian Inference (4)
Univariate meta−analysis (8)
Calibration
Brier Score
0.20
0.15
0.10
0.05
0.00
LIT: Wells (9)
LIT: Modified Wells (10)
LIT: Gagne (4)
LIT: Hamilton (7)
LIT: Oudega (8)
IPD: LRM (4)
IPD: PMLE (6)
Univariate meta−analysis (4)
Multivariate meta−analysis (4)
Bayesian Inference (4)
Univariate meta−analyis (8)

External Knowledge: heterogeneous prediction models

- Robust aggregation approaches
  - Model Averaging
  - Stacked Regressions
- Explicit summary model
  - Estimate averaged predictions
  - Expand regression coefficients



Discrimination
AUC
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
LIT: Wells (9)
LIT: Modified Wells (10)
LIT: Gagne (4)
LIT: Hamilton (7)
LIT: Oudega (8)
IPD: LRM (15)
IPD: PMLE (6)
Model Averaging (15)
Stacked Regressons (15)
Calibration
Brier Score
0.20
0.15
0.10
0.05
0.00
LIT: Wells (9)
LIT: Modified Wells (10)
LIT: Gagne (4)
LIT: Hamilton (7)
LIT: Oudega (8)
IPD: LRM (15)
IPD: PMLE (6)
Model Averaging (15)
Stacked Regressons (15)


Discrimination
AUC
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
LIT: Wells (9)
LIT: Modified Wells (10)
LIT: Gagne (4)
LIT: Hamilton (7)
LIT: Oudega (8)
IPD: LRM (15)
IPD: PMLE (6)
Model Averaging (15)
Stacked Regressons (15)
Calibration
Brier Score
0.00
0.05
0.10
0.15
0.20
LIT: Wells (9)
LIT: Modified Wells (10)
LIT: Gagne (4)
LIT: Hamilton (7)
LIT: Oudega (8)
IPD: LRM (15)
IPD: PMLE (6)
Model Averaging (15)
Stacked Regressons (15)


FORBIDDEN PLANET
42

Strengths and Weaknesses

- Strengths
  - Aggregation usually improves performance
  - Abundance of external knowledge
  - Straightforward implementation of approaches
  - Explicit aggregated models (no black boxes)
- Weaknesses
  - Heterogeneity of external knowledge
  - Performance gain not always very large
  - Additional efforts required during derivation phase

Conclusion

Promising Results

- Simulation studies & Applications

Potential Implementations

- Derivation of robust prediction models using small samples
- Update existing models with new evidence
- Unify disparate models from literature

Future (**Birmingham**)

- Prediction modeling with multiple IPDs
- Publication bias