PREDICTION MODELS

Leonard Leibovici

Seville, Oct 2016

Development and validation of a prediction rule

- What are the aims? Do we need such a prediction rule?
- Derivation group
- How to derive the rule: which model to use
- Internal validation
- External validation
- …a few examples
- Conclusions

What are the aims? Do we need such a prediction rule?

- A perceived problem in decision making.
- For local or universal use?
- A quantifiable problem:
  - How often is a wrong decision made?
  - What measures are disturbed and to what extent?
  - How are patients' outcomes affected?
- Quantify and provide baseline data: for comparison (historical comparisons are weak but might be important to understand trends).
- Examples:
  - Percentage of true positive blood cultures 3-5% (emergency wards); 12-15% (departments of medicine).
  - Pneumonia: huge variability in practice.

What are our needs?

- We should consider what is the purpose of the prediction model.
- To turn it into a true decision aid we need a target function to optimize: a decision analytic approach.

- Examples:
- Bacteremia: a large enough group of patients with almost no true positive blood cultures (i.e. high specificity and good calibration).
- Pneumonia:
  - Define a group with a very low probability of a bad outcome and thus can be managed as outpatients.
  - Define a group with a high probability of a bad outcome (an outcome amenable to intervention) that should be admitted to the ICU.

Derivation group

- Protocol: define:
  - Population of interest (inclusion and exclusion criteria, how detected)
  - Candidate predictors
    - Do they fit the clinical workflow?
    - Biological/clinical plausibility?
  - Outcomes
- Prospective vs retrospective collection of data:
  - Prospective is always preferable.
  - Watch out for external validity.
  - Rare outcomes might necessitate retrospective collection.

Choice of model: Self-learning algorithms

- ‘Black boxes’ that marry a combination of predictors to an outcome.
- Demand large and quite complete databases.
  - Research data bases are usually small and incomplete.
  - Administrative databases are large but were not built for research.
- High connectivity (many degrees of freedom) leads to over-matching.
- Varying (usually low) insight into mechanisms.

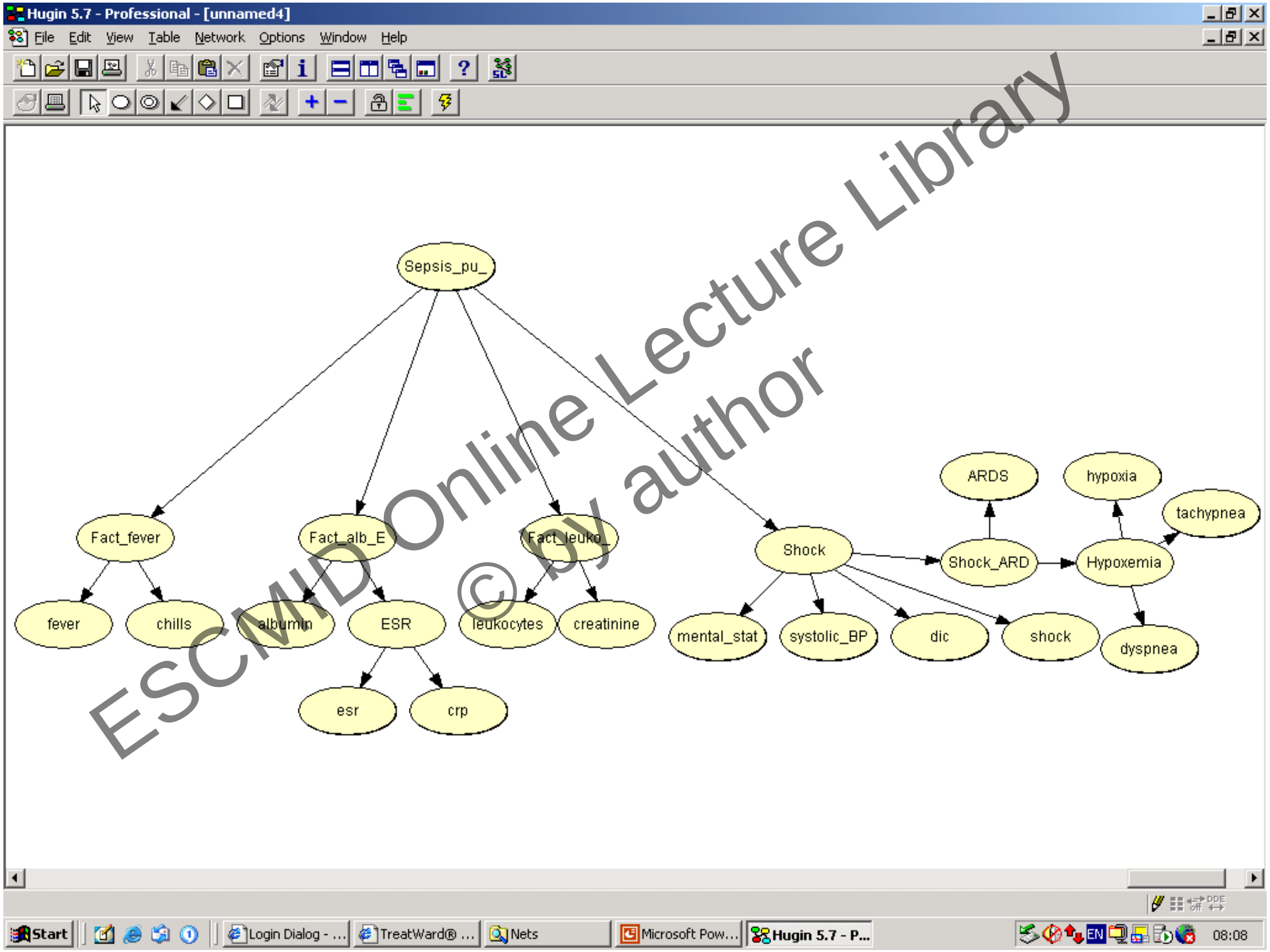
Choice of model: Statistical models, usually logistic regression

- Risk of overfitting.
- Needs statistical know-how but also an excellent grasp of the biological/clinical domain.
- Difficult to update.
- Lacks several of the advantages of causal models.

- But overall we have the most experience with this kind of modelling.

Causal models - advantages

- Modeling of complex situations
- Allows combination of knowledge and data, and of data from different sources
- Allows explicit differentiation between local and universal factors, and thus makes calibration easier
- Decision – analytic approaches integral to some systems (influence diagram)
- Missing information handled by the model itself


Hugin 5.7 - Professional - [unnamed4]
File Edit View Table Network Options Window Help
Sepsis_pu_
Fact_fever
Fact_alb_E
Fact_leuko_
Shock
ARDS
hypoxia
tachypnea
Shock_ARD
Hypoxemia
fever
chills
albumin
ESR
leukocytes
creatinine
mental_stat
systolic_BP
dic
shock
dyspnea
esr
crp
ESCMID Online Lecture Library
© by author
Start
Login Dialog - ...
TreatWard® ...
Nets
Microsoft Pow...
Hugin 5.7 - P...
08:08

Validation:

- Calibration: (do n of 100 patients with a risk prediction of *n*% have the outcome across all the range of the predictor?)
- Discrimination: are my predictions good enough for a specific patient?
- Does it do what I need?
- Do I improve a target function by applying the model?
- Do I improve an outcome by applying the model?

For formal ways to assess the performance of prediction models:
Epidemiology. 2010 Jan; 21(1): 128–138.

Internal validation:

- The model is validated in the same database from which it was derived.
- Boot-strapping techniques.
- Split the database into derivation and (internal) validation set from the beginning: in space or in time.

External validation

- Local or universal?
- In another place (multiple places even better) and (by necessity) another time:
  - Calibration: (do n of 100 patients with a risk prediction of *n*% have the outcome across all the range of the predictor?)
  - Discrimination: are my predictions good enough for a specific patient?
  - Does it do what I need?
  - Do I improve a target function by applying the model?
  - Do I improve an outcome by applying the model?

How do I test whether outcomes were improved?

Designs of tests

	Patient randomisation	Cluster randomisation	Before-after
Educative			
Interacts with the medical setting			
Good answer:	Partial answer	Full answer	Full answer?
Removes bias	Full	Partial	Badly
Publication and impact	High	Partial	Low
Cost	High	Lower?	Lower
Patient consent	needed	probably	??
No of participants		highest	

Implementation (in a trial or in practice)

Electronic patient file:

- Draw the data and calculate the score automatically.
  - Take note that data from electronic files do not always have the same meaning as the variables that were collected.
- Implement the results semi-automatically.

If unsuccessful, what went wrong?

- Included risk factors that are relevant only in some locations.
- The baseline incidence of the outcome is very different.
- Doesn’t fit into the workflow.
- Not accepted by the users (for some reasons).

Predicting bacteraemia in validated models—a systematic review (Clin Microbiol Infect 2015; 21: 295)

Inclusion criteria for studies:

- Validated (either internally or externally)
- Studies that were able to define groups with low or high probabilities for bacteraemia (arbitrarily defined as below 3% or above 30%).

- 21 studies were excluded because they did not have any form of validation.
- 15 studies included (total of 59 276 patients).
- 12 performed external validation.
- 7 models were validated in a different hospital
- In 5 the model performed well.

Predicting bacteraemia in validated models—a systematic review (Clin Microbiol Infect 2015; 21: 295)

We contacted the authors of these 5 studies. None of them were implemented in clinical practice.

Risk stratification: independent cohorts: TREAT performance

	Cohort 1			Cohort 2		
	N	**Bacteremia (%)**	**Contamination (%)**	**N**	**Bacteremia (%)**	**Contamination (%)**
Low-risk group	123	3 (2.4)	5 (4.1)	300	4 (1.3)	9 (3.0)
Intermediate-risk group	483	62 (12.8)	12 (2.5)	1139	150 (13.2)	61 (5.4)
High-risk group	184	55 (29.9)	10 (5.4)	285	80 (28.1)	16 (5.6)

TREAT module for predicting bacteremia is not used.

- It was not accepted by the hospital Antibiotic Committee: “The information in blood cultures goes beyond negative/positive. We are not convinced.”
- TREAT uses lab values: blood is drawn for blood culture with the first venipuncture.

Value of severity scales in predicting mortality from community-acquired pneumonia: systematic review

(Thorax. 2010;65:884)

- Included prospective studies that reported mortality at 4-8 weeks in patients with radiographically-confirmed community-acquired pneumonia.
- Test performance was evaluated based on 'higher risk' categories.
- 23 studies involving 22,753 participants (average mortality 7.4%) were retrieved.
- Negative predictive values for mortality were similar among the tests, ranging from 0.94 (CRB-65) to 0.98 (PSI).

A non-infectious example

- CHA2DS2-VASc and CHADS2 are in frequent use.
- The scores function no better than the ones we have reviewed.
- ??

Conclusions:

- We can draw a few helpful guidelines for assembling a clinical prediction model:
  - Ask whether the model in needed and what should it do.
  - Consider validation of an existing model instead of derivating your own.
  - Define carefully your derivation group.
  - Choose your model, but examine causal paths.
  - Test the model in an independent cohort.
  - Test its performance in clinical practice: does it change management or outcomes? Choose your study design.
  - Examine how it fits into the workflow.
  - Integration into electronic patient file.

Questions for you:

- Why these efforts are not more successful?
- Should they be used more often?

Thank you for your attention.