Statistical considerations for the development of diagnostic tests

Olivier Collignon, PhD

European Medicines Agency, Luxembourg Institute of Health

Workshop on Site and Histology – Independent Indications in Oncology
EMA, 14/12/17

Disclaimer

The views expressed are those of the presenter and should not be understood or quoted as being made on behalf of the Luxembourg Institute of Health or of the European Medicines Agency (EMA) or its scientific committees or reflecting the position of the EMA.

Rationale...

- Histology-independent trials often start with a molecular screening phase (basket trials);

- Sub-studies correspond to different target populations, defined by different diseases or sub-types of the same disease;

- Positivity to a given biomarker is often the common denominator between the different sub-studies;

- Accuracy of the biomarkers is the corner stone of these designs, otherwise false positive patients can be enrolled

… but also a Trojan Horse

http://projectbritain.com/calendar/April/trojanhorse.html

What do we need to develop a diagnostic test?

- The objective is to replace an invasive/expensive gold-standard of detection by a diagnostic test (say, positivity of a biomarker)

- Clear gold-standard to differentiate patients
    - Disease diagnosis: biopsy
    - Companion diagnostic : criteria to define treatment benefit ?

- Case-control study of the highest quality
    - Cases (gold-standard +);
    - Controls matching the cases to the best possible (gold-standard -)
    - Baseline characteristics (ex: age, gender, clinical history, etc)
    - Panel of robust biomarkers (controlled measurement errors, within-patient variations);

How to develop a diagnostic test?

_ There might be thousands of candidates in a company pipeline;

_ A diagnostic test can be based on many biomarkers, not necessarily a single one;

_ Keeping irrelevant biomarkers is costly and inefficient;

Usual work plan:

1. assess the univariate predictive performance of each biomarker one by one;
2. select the most performant combination of biomarkers;
3. show that this combination outperforms / complements what is easily predicted using variables collected in routine (e.g. age, gender, clinical characteristics,…)

A diagnostic test is a clinical prediction model

Main steps for building a clinical prediction model:

1. Biomarker selection
2. Combination of the biomarkers and building of the predictive model
3. Cut-off selection to maximize the performances of the tool in terms of sensitivity and specificity
4. Internal/external validation of the model

Often some of these milestones can be achieved simultaneously

1. Biomarker selection

_ There is a core set of statistical methods to guide the biomarker selection:

- univariate screening based on p-value < cut-off
- stepwise logistic regression
- LASSO regression
- ...

Univariate screening and stepwise logistic regression are better avoided (multiplicity issues, a tendency to provide false-positive markers,...)

LASSO tends to be the gold-standard among methodologists

2. Combination of the selected biomarkers

_ Very often the selected biomarkers are combined within a **logistic regression**;

_ The observed values of the biomarkers $x_1, \ldots, x_p$ are given weights $\beta_1, \ldots, \beta_p$ to build the probability of having a positive test, given by the formula:

$$\pi(x_1, \ldots, x_p) = P(being\ a\ case | biomarkers\ levels)$$

$$= \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}$$

3. Cut-off selection

The diagnostic test is declared positive for a given patient if, having observed his biomarker measurements, the probability π is greater than a given cut-off α (e.g. 50%, 80%).

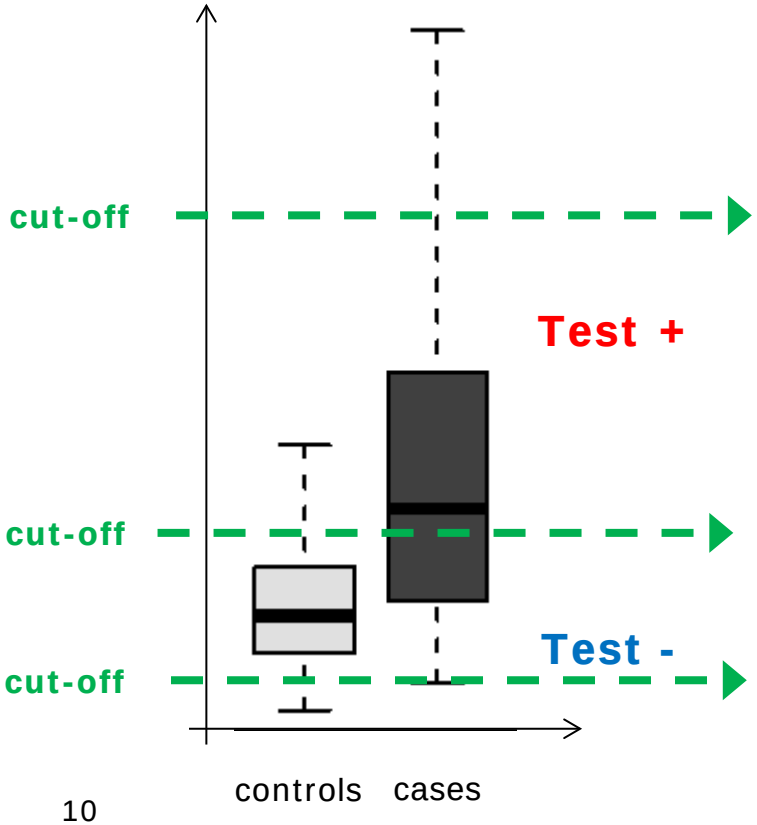
Sensitivity $= P(test+ | being\ a\ case) = P(\pi > \alpha\ | being\ a\ case)$

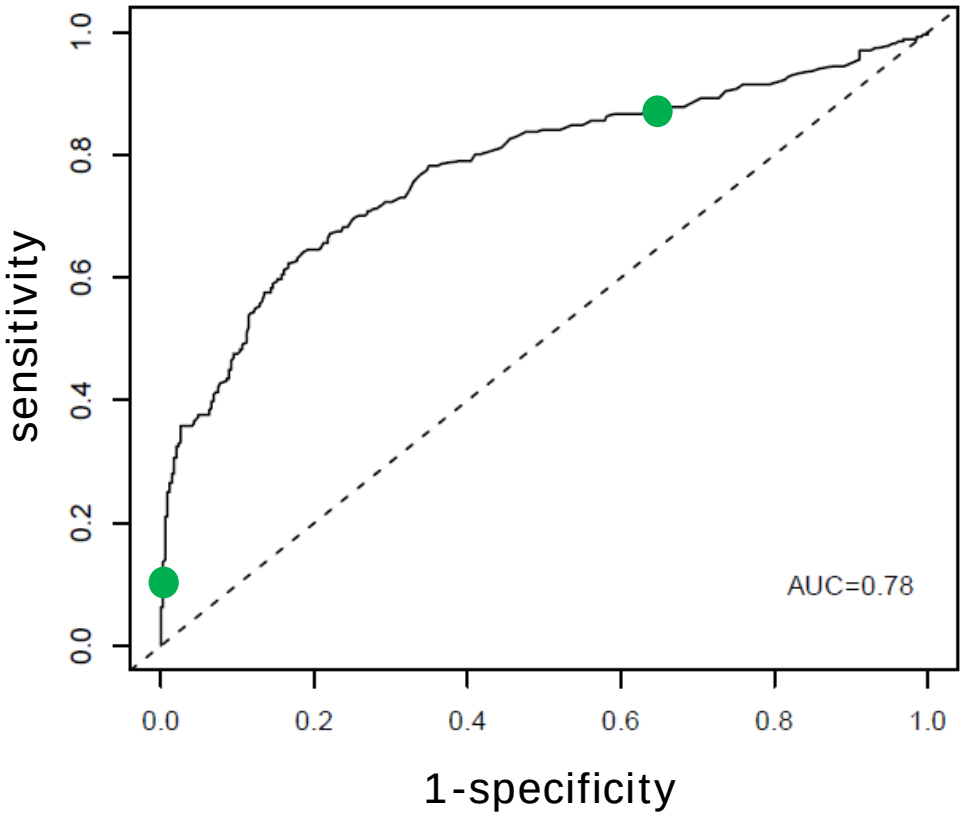
Specificity $= P(test- | being\ a\ control) = P(\pi \leq \alpha\ | being\ a\ case)$

Each cut-off defining positivity of the test leads to a certain sensitivity and specificity of the method;

3. Cut-off selection - ctd

Varying the cut-off enables the construction of the Receiver Operating Characteristics (ROC) curve by plotting the point (1-specificity,sensitivity) for each cut-off considered;
The AUC is an ideal measure of performance because it is cut-off independent.





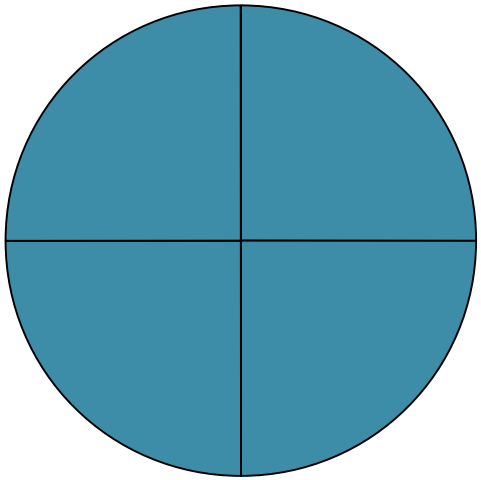
3. Cut-off selection - ctd

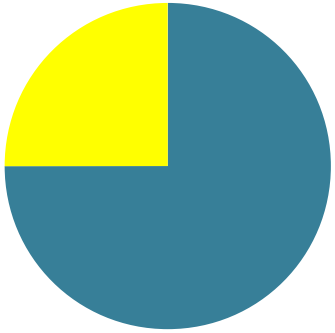
Often the optimal cut-off is obtained by maximizing a compromise between sensitivity and specificity, for example:

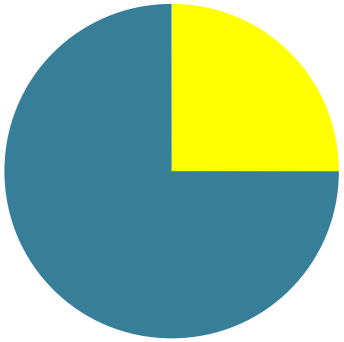
- sensitivity+specificity-1 ("*Youden Index*")
- $(\text{sensitivity}-1)^2+(\text{specificity}-1)^2$
- maximize sensitivity when specificity is fixed at 95%

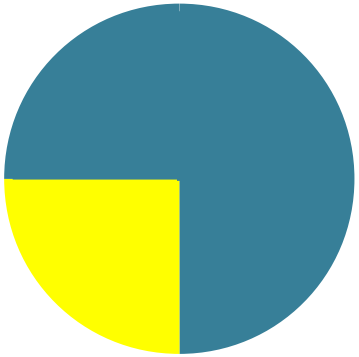
4. Internal validation

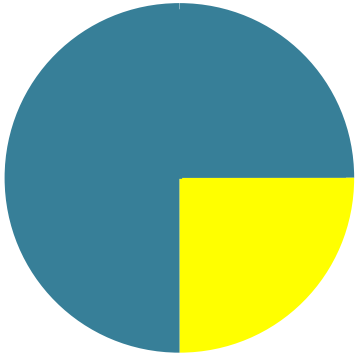
Ex: cross-validation

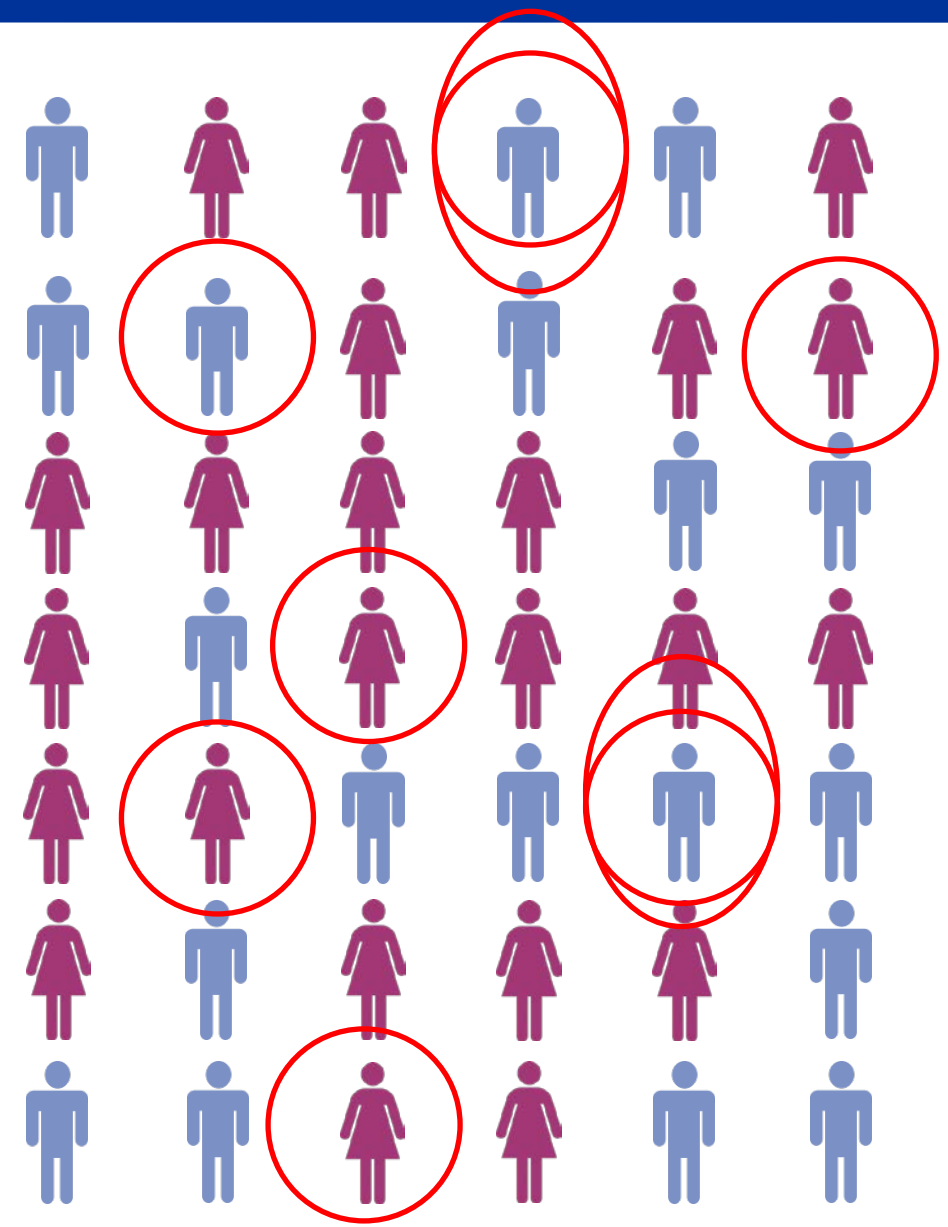






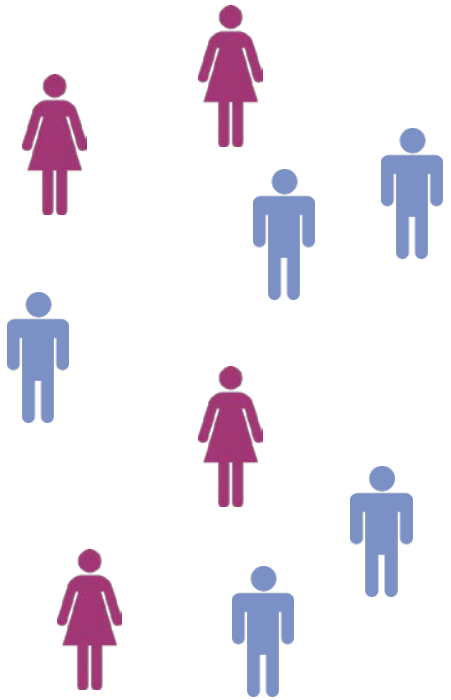






Bootstrap internal validation

Bootstrap sample



External validation

Another dataset with the same characteristics (e.g. inclusion criteria, protocol) should be collected in order to evaluate the sensitivity and specificity of the selected cut-off in a independent setting.

Ideally the characteristics can be further widened (e.g. less restrictive inclusion criteria) in order to evaluate the robustness of the sensitivity and specificity of the test to small variations in the population of interest.

References

1. Steyerberg EW. Clinical Prediction Models: A Practical Approach To Development, Validation, And Updating. New York: Springer; 2009.

2. Harrell FE. Regression Modeling Strategies: With Applications To Linear Models, Logistic Regression, and Survival Analysis. New York: Springer; 2001.

3. Collignon, Olivier. "Methodological issues in the design of a rheumatoid arthritis activity score and its cut-offs." *Clinical epidemiology* 6 (2014): 221.