**Statistical approach to development and validation of patient-reported outcomes tools for individual decision-making**

**Author's name(s):** Regnault, A*; Gilet, H; Meunier, J; Arnould, B  
**Affiliation(s):** Mapi Values, 27 rue de la Villette 69003 LYON  
**Email:** antoine.regnault@mapivalues.com  
**Phone:** +33(0)4 72 13 59 79 ; **Fax:** +33(0) 72 13 51 40  
**Corresponding author:** A Regnault

**Keywords:** Individual decision-making; Development; Validation

**Topic area of the submission:** scale validation

**Introduction**

The methods pertaining to psychometrics, either modern or classical test theory, that are usually used to develop and validate Patient-Reported Outcomes (PRO) instruments are appropriate when the purpose of the application of the questionnaire is to measure a construct in a group. However, PRO instruments can be used with another purpose: they can contribute to patient management in clinical practice by supporting decisions for a given individual. This change in context of use implies a deep change in paradigm[1] that impacts the whole development and validation process including the statistical method to be used.

The objective of this presentation is to describe the statistical method to be applied to build and validate PRO tools for individual decision-making, using illustrative examples based on published studies.

**Statistical approach to PRO tool for individual decision-making**

PRO tools could contribute to the answer to many possible questions, such as: Should this patient be considered for the diagnosis of disease A? (Screening tool); Does this patient suffer from disease A? (diagnosis-aid tool); Is this treatment indicated to this patient? (prescription-aid tool); Is the treatment taken by this patient the right one? (treatment adaptation tool). For all of these situations, the final purpose of the PRO instrument application is not to assess a continuous latent variable but to be able to use patients’ responses to separate a discrete number of patient groups, corresponding to the different possible answers to the question of interest.

The approach of choice to develop and validate PRO tools for individual decision-making is to collect the PRO data for patients for whom the response to the question of interest is known and to optimize and validate the tool on this sample that includes all the information needed
for all patients. In this approach, the definition of the gold standard (i.e. the criterion used to identify in the study sample the groups to be separated) is critical since all the process will be designed to replicate these groups.

**Construction of PRO tool for individual decision-making**

The first step of the statistical approach for the development of PRO tools for decision-making is the construction of a discriminant model. The discriminant model is a function of the items of the PRO tool that enables the separation of the patient groups as identified by the gold standard. From a statistical method perspective, the construction of the discriminant model involves classification analysis techniques: linear discriminant analysis, logistic regression, classification trees, etc. Only a subset of the various existing classification methods is appropriate in this context. Indeed, they should allow the production of discriminant models that are easy to implement (no complex scoring is possible in routine medical practice), transparent (the user should be able to understand how the different items contribute to the decision) and still captures the complexity of the PRO data.

Then, a decision rule should be derived from the discriminant model. Depending on the method used to define the discriminant model, this step can be straightforward (e.g. classification trees) or necessitate an additional step (defining a threshold on a score). In this later case, the classification rule is generally defined by minimizing a misclassification function.

For example, classification and regression tree (CART) was applied to the data of the DN4 tool, a tool designed to separate patients with two different type of pain (nociceptive vs. neuropathic): the resulting discriminant model included only 3 variables and efficiently separated the patient with the different types of pain[2].

**Validation of PRO tool for individual decision-making**

The main objective of the PRO tools in the decision-aid framework is to enable good decisions to be made. Hence, the validation of PRO decision-aid tools focuses on properties related to this objective. The first property of importance in this purpose is the discriminant accuracy of the discriminant model. The discriminant accuracy characterizes the extent to which the discriminant model is able to separate the different patient groups. The classical estimation of discriminant accuracy is the area under a Receiver-Operating Characteristic (ROC) curves.
The second validation criterion is the ability of the decision rule to correctly classify the patients. Several indicators are possible, and their relative importance depends on the context: overall correct classification rate, sensitivity, specificity.

A final approach that can be of great interest in the validation of PRO tools for decision-making is to investigate the ability of the tool to predict a future outcome of importance related to the decision made.

For example, the 12-item score of the ADherence Evaluation of OSteoporosis (ADEOS) tool was build to separate compliant and non-compliant patients (gold standard to define compliant patient: score of 4 to Morsiky Medication Adherence Scale), and showed good predictive accuracy (area under ROC curve: 0.84). It also demonstrated good ability to predict patients who were at risk of stopping taking their treatment in the following 9 months: the relative risk of stopping the treatment for patients with a score lower than 8 compared to those who had a score higher than 12 was 1.69 [3].

**Control of learning bias**

Generally, the discriminant model is obtained by maximizing the separation of the same groups that are used for the assessment of discriminant accuracy. The estimation of discriminant accuracy on the sample on which the discriminant model has been obtained is therefore over-estimated. Several methods exist to control for this learning bias while using a unique sample for development and validation: split-sample validation, cross-validation, or bootstrap validation.

For example, a study was conducted to separate three upper gastro-intestinal disorders – gastro-esophageal reflux disease, ulcer-like dyspepsia and dysmotility-like dyspepsia – using the symptoms perceived by the patient. The tool developed using Partial-Least Square Discriminant Analysis (PLS-DA) allowed the correct classification in one of the 3 groups of 54.7% of the patients of the construction sample and this performance did not deteriorate on an independent test sample (correct classification rate on test sample: 56.8%). This result demonstrated the robustness of the PLS-DA results[4].

**PRO measurement and individual decision in clinical practice**

PRO instruments designed to measure a continuous latent construct are generally not appropriate to support decision in clinical practice. However, they can be used for individual patients at certain conditions. The first critical prerequisite is that they address a clear question relevant to clinical practice and, ideally, that their development process has been adapted to
this final purpose. Second, they have to meet particular criteria for some psychometric properties: very high reliability (for the measurement error of an individual measure to be very low); good discriminant properties between clinically relevant groups. Finally, reference values should be developed for the interpretation of the measure to be direct and for subsequent decision to be made easily.

For instance, the Disability RElated to COPD Tool (DIRECT) is a 10-item measurement instrument designed to help physicians to consider the disability of their COPD patients. It was designed specifically from a targeted clinical question, relevant to clinician. It showed very high reliability, and ability to separate between patients of different clinical severity [5]. Finally, guidelines for the interpretation of the DIRECT score are being prepared.

**Conclusion**

In order to address the specific issues related to the individual decision context, the development and validation of PRO tools to be used as decision-aid tools in clinical practice require the application of specific statistical methods, which differ from those classically used in the clinical research context.

**References:**

[1] Arnould,B. Patient-reported outcomes and clinical practice. From measurement instruments to decision tools: much more than simple change in format. PRO Newsletter, 2006, 36, 21-24


