Summary

At the beginning of this thesis, we propose Poisson processes for modelling patient recruitment in enrichment clinical trials. These processes consider the prevalence of biomarker positive patients in the unselected patient population, the rate of patients submitted to the screening and marker-test procedures, as well as marker-test characteristics. We derive the distributions of the recruitment time analytically, which represent a significant cost component (e.g. personnel, capitalization and infrastructure costs). We suggest models for study cost evaluation that consider the costs associated with screening, marker-test, and patient care during the trial. To evaluate the costs, we derive the distribution of the number of patients passing through the screening and test phases. The impact of false positive patients on the study's power is investigated; a marker-test with low specificity allows enrollment of some false positive patients which may decrease the study's effectiveness. We propose a Bayesian approach for progressively updating components like marker prevalence, recruitment rate, and the time remaining to study completion . In summary, the smaller the marker prevalence, the longer the recruitment time and the higher the study costs. The suggested models both quantify these important factors of enrichment studies and have generalized properties through parametrization (e.g. sensitivity=specificity=prevalence=1).

This thesis goes on in its second section to deal with binary classification problems in high-dimensional settings where the two classes have different class importance and it is relevant to guarantee a minimum pre-determined value of sensitivity (true classification in the most important class). We propose a new approach to entirely control the sensitivity in building binary classification rules based on the optimization of loss functions of binary classification, under the constraint that the sensitivity belongs to a pre-determined interval of admissible values such as $[90, 100]\%$. This strategy is illustrated by considering both the optimization of the likelihood function in the case of logistic regression and the Youden-index formulated as the objective function of binary classification. A L_1-norm penalty of the model parameters is used to select relevant predictors, and we derive conditions for discarding irrelevant predictors. This new approach is designed to provide the best classifier with the pre-determined admissible sensitivity value. The analyses of data related to endometriosis trials verify this claim as presented in the results section.