Article:
Mierzynska, J, Piccinin, C, Pe, M et al. (10 more authors) (2019) Prognostic value of patient-reported outcomes from international randomised clinical trials on cancer: a systematic review. The Lancet Oncology, 20 (12). e685-e698. ISSN 1470-2045

https://doi.org/10.1016/S1470-2045(19)30656-4

Table 1: Main methodological criteria used in the evaluation of prognostic factor analysis

Criteria	Description
Sample size	The number of patients involved in a RCT should be sufficiently large to capture enough events and therefore give reliable prognostic information. Given the implications for study precision and power, sample size should be explicitly reported in publications.
Missing data	Missing data is usually present in PRO data and complicates the generalizability of PRO results to the trial's population. Therefore, reporting the amount of missing data is critical. Additionally, as different kinds of missing data may impact PRO results, reasons why PRO data are missing should be reported. To reduce their negative impact, several guidelines and analysis strategies have been developed. For example, instead of deleting entire patient records, imputation techniques can be used.
A priori selections of PROs predictors	As an a priori selection prevents the risk of selecting potential prognostic factors by chance, model overfitting and type I error, a thorough knowledge of the subject through literature reviews, preclinical data, clinical expertise, etc. should guide the selection of predictors. A description of the process is important to ensure the reliability and generalizability of the results. This selection is even more important with multidimensional and broad concepts such as PROs or HRQoL.
Interaction	Taking into account the interaction between potential prognostic factors engenders additional analyses that could increase the risk of biased results and is therefore not advisable.
Type of variables	PROs scores can be categorical or continuous. The latter is preferred to extract the maximum amount of information. However, when data are reported as categories, it is advised to define them a priori.
Model building strategy	The model chosen should aim to verify if the additional PRO prognostic information helps to better predict the length of OS compared to clinical factors only. Univariable and/or multivariable cox PH models are the main models used in this type of analysis. Different strategies are possible: prognostic factors can be selected from a set of parameters using different methods, e.g., forward, backward, or stepwise variable selection. Forced inclusion of clinical factors may also be used to reinforce the prognostic accuracy by investigating the extent to which PROs factors add prognostic value.
Hypothesis	The formulation of a hypothesis is important in order to focus on a specific endpoint and minimize the risk of bias in analyzing and interpreting results.
Verification of model assumptions	Whether univariable or multivariable analyses are used, the model assumptions should be verified before the formal analysis to ensure that the correct method is performed and appropriate conclusions are drawn.
Quantifying predictive accuracy	Prognostic accuracy refers to the assessment of the additional prognostic value of PROs on top of the clinical factors. PROs are only relevant as prognostic factor if the prognostic significance for OS is statistically but also clinically improved. Different measures are used for this assessment: discrimination C-index, Schemper residuals, Nagelkerke's R^2-coefficient, Likelihood test and PLS regression.
Model validation	Model validation provides a solution in order to avoid overfitting. It helps determine the degree to which multicollinearity might affect the analysis. While internal validation is often sufficient (bootstrap sampling), the most accurate test involves external validation.