

Criteria	Rationale
1. Clinical/conceptual relationship with the outcome of interest	A logical theory must explain the association between the factor and the outcome. Begin with conceptual model informed by research and experience; does not require a direct causal relationship
2. Empirical association with the outcome of interest	A statistical association to confirm the conceptual relationship
3. Variation in prevalence of the factor across the measured entities	If there is no variation in prevalence across providers being measured, it will not bias performance results
4. Not confounded with quality of care – risk factors should: <ul style="list-style-type: none"> • be present at the start of care and • not represent the quality of care provided (e.g., treatments, interventions, expertise of staff) 	<p>Trying to isolate effects of the provider – quality of care</p> <p>Ensures not a result of care provided</p> <p>Although these could explain variation in outcome, trying to isolate differences in performance due to differences in the care provided</p>
5. Resistant to manipulation or gaming – generally, a diagnosis or assessment data (e.g., functional status score) is considered less susceptible to manipulation than a clinical procedure or treatment (e.g., physical therapy)	Ensures validity of performance score as representing quality of care (vs. for example, up coding)
6. Accurate data that can be reliably and feasibly captured at a reasonable cost	Data and resource limitations often represent a practical constraint to what factors are included in risk models.
7. Contribution of unique variation in the outcome (i.e., not redundant or highly correlated with another risk factor)	Prevent over-fitting and unstable estimates, or coefficients that appear to be in the wrong direction, reduce data collection burden
Potentially , improvement of the risk model (e.g., risk model metrics of discrimination – i.e., sensitivity/specificity, calibration) and sustained with cross validation	<p>Change in R-squared or C-statistic may not be significant, but calibration at different deciles of risk might improve</p> <p>May not appear to be a big change but could represent meaningful differences in terms of the outcome (e.g., lives, dollars)</p> <p>Order of entry into a model may influence this result</p>
Potentially , face validity and acceptability	Some factors may not be indicated empirically, but could improve acceptability – need to weigh against negative impact on model, feasibility and burden of data collection

