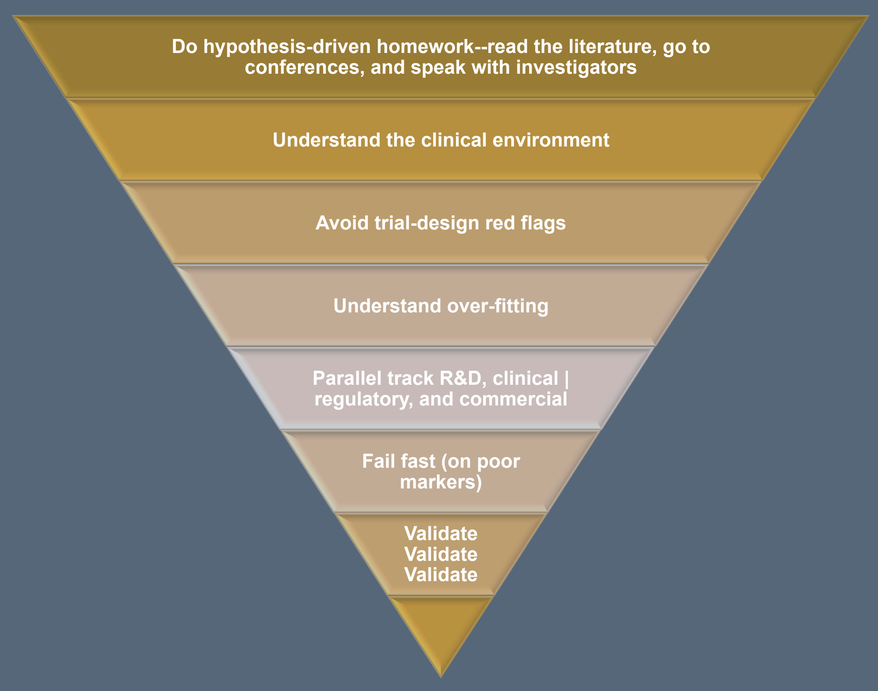
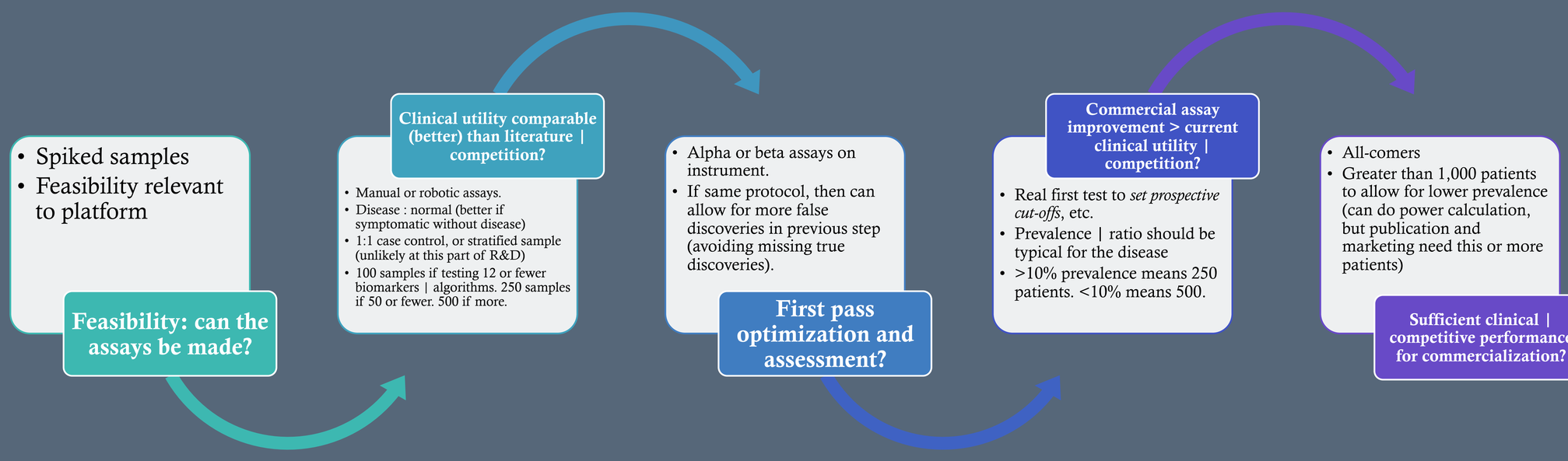
Biomarker Discovery and Validation



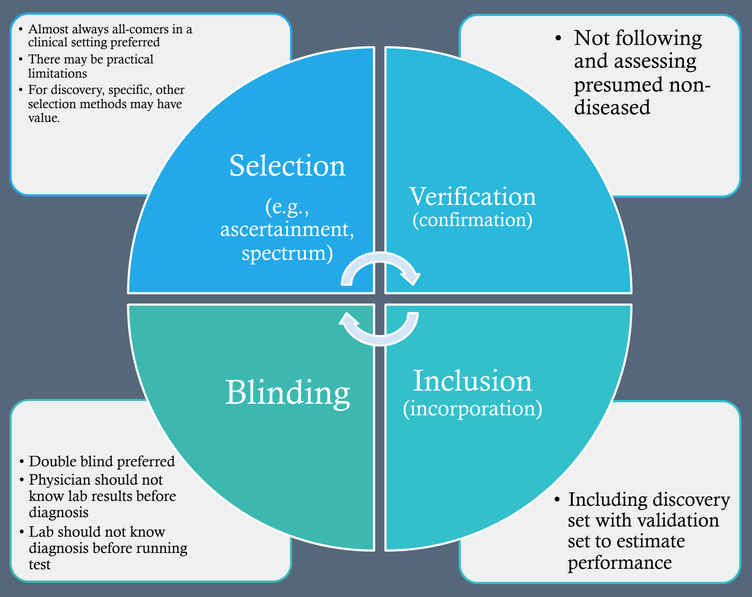
Important lessons on proportions

- Don't use less than 250 patients even when assessing only a few markers
- Start to beware retrospective individual marker discovery at 50 potential markers, in the context above
- For multi-marker indices, beware starting at 25 potential markers
- When prevalence below 12%, then use more than 1,000 patients
- If using 500 to 1,000 patients with prevalence greater than 12%, relatively good even up to 100 markers.

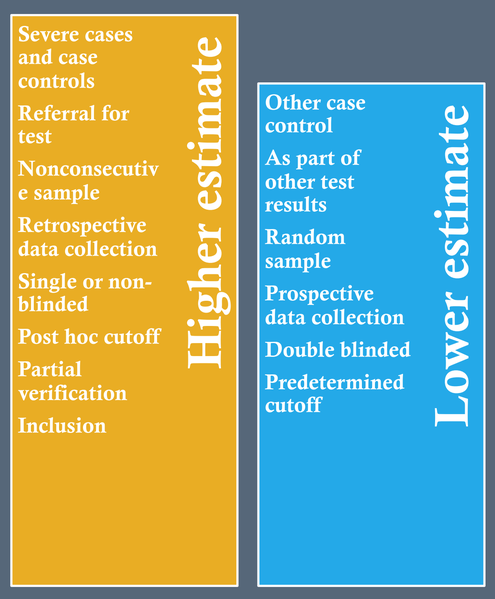
Concurrent assay development and clinical testing



Big Four Biases



Biased Design Effects on Relative Estimate of Diagnostic Performance



False Discovery Assessments

Bonferroni correction (most conservative)

- Divide the desired p value (probability of true discovery) by the number of biomarkers or algorithms tested.
- This establishes the new p value that any biomarker or algorithm must pass.

False discovery rate

- Similar to Bonferroni for assessment of the biomarker | algorithm with the best p value.
- For subsequent, the desired p value is divided by the number of biomarkers | algorithms remaining to be assessed (i.e., the correction gets easier if some biomarkers | algorithms pass)

Further notes on the discovery simulation for estimates of samples sizes

Degrees of freedom can dramatically affect retrospective biomarker analysis.

- Simulations run tested whether as the number of markers investigated increases, and either the prevalence, or number of patients decrease, **the higher the risk for perceived but random positive results in *marker mining*.**
- In order to assess the likely outcome of this effect within the realm of marker mining, an experiment was run using random data sets, and varying the quantities of the three variables just listed.

False AUCs (c-statistics) can be quite high

- Average experimental AUC for *random* single markers was 0.62, with the highest a whopping 0.97
- Average experimental AUC for *random* multi-marker indices was 0.65, with the highest **1.00**

Sample size (number of patients), prevalence, and number of markers mined are important variables to assess against random results

- The major danger zone appear to be characterized by patient sizes less than 250 (for essentially *all prevalence values,* and even if mining *only a few markers*)
- Additionally, *when mining 25* or more markers, a prevalence below 12% raises concerns, even with patient *sizes up to 1,000*
- The converse of this seems to indicate that *patient sizes of 500 to 1,000* appear to obviate positive random results *even* when mining 100 markers as long as the *prevalence* is greater than 12%

Source: Canadian Medical Association Journal, proprietary simulations