A NEWLY DEVELOPED POWER PROCEDURE FOR SAMPLE SIZE PLANNING IN COA VALIDATION STUDIES

Daniel Serrano, PhD, Charles Iaconangelo, PhD & MS, & Jarjieh Fang, MPH

Patient Centered Outcomes

pharmerit international

Background

A power procedure for validation study planning does not exist.
Sample size planning in PRO has historically relied on rules of thumb:
- "10 people per item"

This disregards:
- the number of response categories
- the number of domains measured by the PRO
- Sample size evidence from EFA/CFA simulations indicating minimum n=300

When power analyses have been used:
- Power has focused on classical methods or effect detection
- This requires untestable assumptions about scores (e.g. unidimensionality and normality)

Objective

Develop a procedure for powering domain specification
Domain Specification analyses have the greatest sample size requirement of all validation analyses
This protects downstream analyses from errors emerging from underpowering the first analysis of validation: domain specification

What to Power for in Validation and Why

Powering for classical methods or effect detection relies on unverifiable assumptions about the PRO scoring procedure.

- Generally assumes a single score disregarding either the conceptual framework or empirical domain specification

However, empirical scoring evidence emerges as a result of domain specification analyses

Such analyses are most efficiently executed using modern methods

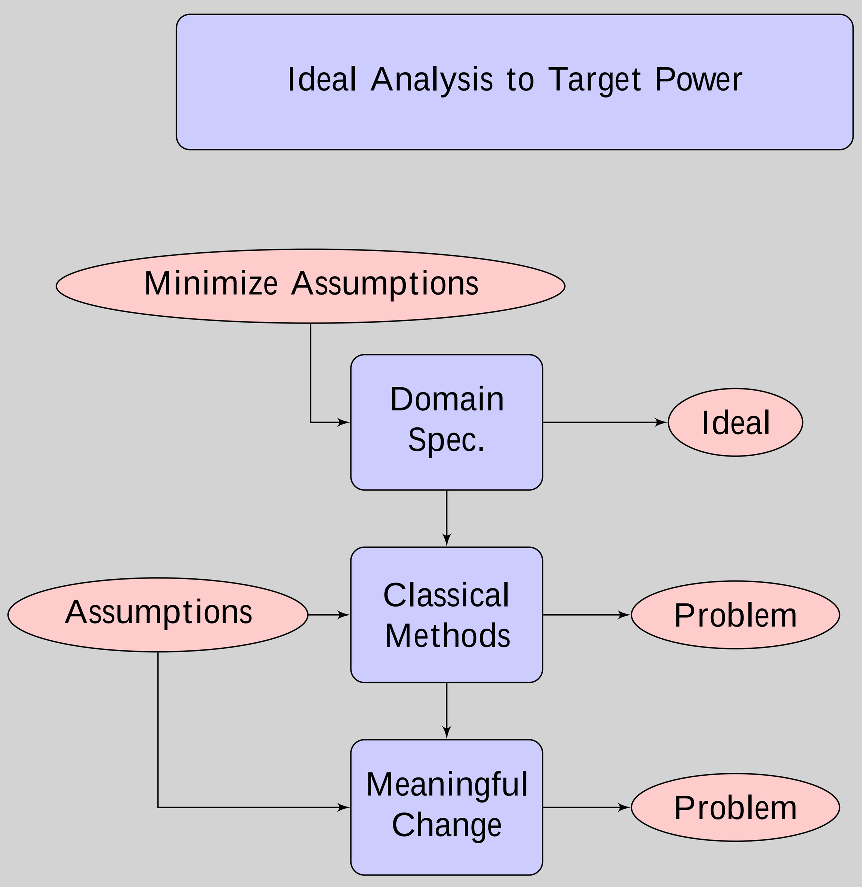

Figure: Power Computations

Power Computation

How to power domain specification analyses?
- Model fit statistics
- Powers for ability to detect appropriate domain specification
- Equal to powering for content validity

What model fit statistic to use?
- Root Mean Squared Error of Approximation (RMSEA)
- RMSEA follows $^{2}_{DF;}$ distribution
- Several RMSEA variants have been developed for Full Information Models (Item EFA and IRT)
- These include $M2$, $M2$, and $C2$

How to estimate power for RMSEA
- Compute DF for planned model
  - J Items with K response categories
  - There are $M = J + J(K \quad 1)$ estimable item parameters
  - In the case of EFA, with D domains/factors, one deducts the $\frac{D(D \quad 1)}{2}$ inter-factor correlations from M
  - The estimated parameters, M, are subtracted from the statistic-specific moments
  - Each statistic ($M2$, $M2$, and $C2$) computes moments differently
  - Statistic-specific DF are given in Equation 1

$$DF_{M2} = J \quad (K \quad 1) + \frac{J}{2} \quad (K \quad 1)^2 \quad M$$
$$DF_{M2} = J + \frac{J}{2} \quad M \qquad (1)$$
$$DF_{C2} = J \quad (K \quad 1) + \frac{J}{2} \quad M:$$

One then specifies the non-centrality parameters ( ) for the $^{2}_{DF;}$ distribution
This corresponds to the hypothesis tested.

Power Computation Contd.

Three hypotheses exist:
- Exact fit: Under the null, $_0 = 0$; under the alternative, $= {}_a$
- Close fit: $_0 < {}_a$
- Not close fit: $_a < {}_0$

Power is the area between $^{2}_{DF;\ 0}$ and $^{2}_{DF;\ a}$ for each hypothesis

Simulation Study

Conducted to test alignment between theoretical and empirical
Misfit induced by generating data from bifactor model (Table 1) and fitting a unidimensional model

These two models often compete in PRO development

Power to select correct model has direct bearing on scoring algorithms.

Item parameters selected to produce range of power between $n = 100$ and $n = 500$ achieving power of $0{:}8$

$n = 100$ to $n = 500$ reflects PRO validation sample sizes

$r = 1000$ replications per simulation condition

Table: 9-item Bifactor Generating Parameters

	slopes			intercepts			
Item	g	$r1$	$r2$	C_1	C_2	C_3	C_4
1	2.0	0.5	0.0	2	1	-1	-2
2	2.0	0.1	0.0	2	1	-1	-2
3	2.0	-0.5	0.0	2	1	-1	-2
4	2.0	-0.1	0.0	2	1	-1	-2
5	2.0	1.0	0.0	2	1	-1	-2
6	2.0	0.0	0.1	2	1	-1	-2
7	2.0	0.0	-0.1	2	1	-1	-2
8	2.0	0.0	-0.5	2	1	-1	-2
9	2.0	0.0	0.5	2	1	-1	-2

Simulation Results

Theoretical power calculation accurately reflects empirical power

Simulation results for 9-item condition presented

Results identical for 20-item condition

Regardless of sample size:
- Theoretical power accurately reflected empirical power

Power procedure validated

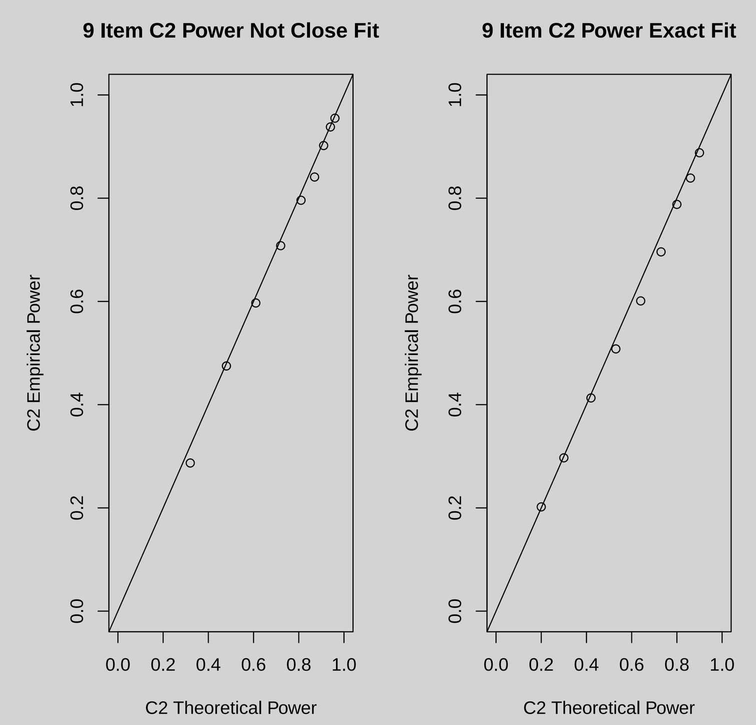

Figure: C2 Theoretical Vs. Empirical Power, 9 items

What matters? Items, response categories, or both?

- In the case of $M2$, adding response categories reduces power
- Adding response categories adds intercepts to estimate
- Adding response categories increases the moments the model must explain

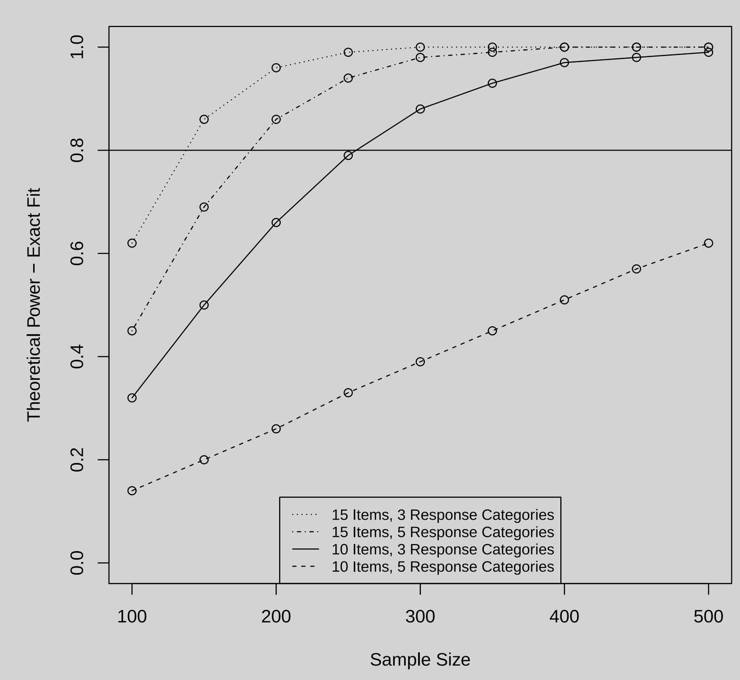

Figure: Items, Response Categories, and Power

Conclusions

Procedure accurately powers for Domain specification/content validity
Procedure minimizes assumptions made:
- Assume Items (J)
- Assume Response Categories (K)
- Assume Domain Specification { can be based upon Revised Conceptual Framework

Procedure can be used early in item generation to understand power implications
A power GUI will be freely available at Pharmerit.com