Design and Analysis Considerations in Research with Small Samples

Rick H. Hoyle
Department of Psychology & Neuroscience
Duke University

NAS Committee on National Statistics
Improving Health Research for Small Populations: A Workshop
January 18-19, 2018

Topics

- When are analyses informative?
- What do we mean by "small"?
- Finite population correction
- Research strategies that address small-sample concerns
- Multivariate models

Analyses are informative when they . . .

1. address the question that motivated the research . . .
   or, address a narrower or more preliminary question for which an analytic strategy can be justified given *N*;
2. use data that satisfy the assumptions of the analytic strategy;
3. are sufficiently powered to detect meaningful effects . . .
   or, reveal descriptively promising patterns in the data so that a new, more focused and informative study can be run;
4. produce results likely to generalize to the target population.

What do we mean by "small"?

- Sample size is small when …
  - estimates and tests would be unduly influenced by a small number of cases;
  - it falls at or below the minimum required for valid estimates of parameters and/or standard errors;
  - estimation sometimes results in nonconvergence or produces problematic parameter estimates;
  - statistical tests are insufficiently powered to detect meaningful effects.

Small *and* constrained

- small sample solutions are for circumstances when samples are small and <u>constrained</u>
  - population of cases is small
  - reaching cases requires substantial resources or is otherwise infeasible
- proposed solutions are not for circumstances in which sample size is not constrained; in such cases, increasing sample size is the preferred solution
  - the compromises required when using small data are not justified when sample size is small but could be increased with reasonable investment of additional resources

When sample size is constrained by population size

- sampling fraction, $f = \frac{n}{N}$, where *n* = sample size and *N* = population size
- $f = 1$ = census (i.e., no sampling error)
- as *f* approaches 1, standard error is adjusted downward to reflect reduction in sampling error due to increasingly large proportion of population in sample
- as *f* approaches 0, tests mirror those when population is assumed to be an infinitely large superpopulation
- when $f > .05$, power of statistical test can be improved through use of the finite population correction factor

$$FPC = \sqrt{\frac{N-n}{N-1}}$$

Finite population correction factor (Cochran, 1977)

$$FPC = \sqrt{\frac{N - n}{N - 1}}$$

- applied to standard error for tests of parameter estimates
- example, $\sigma_{\mathrm{M}} = 10$, $N = 200$, n varies; $\sigma_{\mathrm{M}*} = \mathrm{FPC}(\sigma_{\mathrm{M}})$

n	*f*	FPC	$\sigma_{\mathrm{M}*}$
175	.875	.354	3.54
150	.750	.501	5.01
125	.625	.614	6.14
100	.500	.709	7.09
75	.325	.793	7.93
50	.250	.868	8.68
25	.125	.938	9.38
10	.050	.977	9.77

Finite population correction factor

- can be used for study planning when working with finite population
  - determine required sample size, n_r, if assuming infinite population sampled with replacement
  - derive sample size adjusted for planned use of FPC, n_a

$$n_a = \frac{n_r}{1 + \frac{(n_r - 1)}{N}}$$

N	n_r	n_a	n_a/n_r
200	150	86	.57
200	125	77	.62
200	100	69	.69
200	75	55	.73
200	50	41	.82

Finite population correction factor

- assumes random sampling without replacement
- accounts for reduction in sampling error as *f* increases toward 1.0
- allows inference about state of population at that point in time; not prediction of state of other populations or state of current population at a later point in time

Research strategies for addressing small-sample concerns

$$T = \frac{parameter\ estimate}{standard\ error}$$

- options
  - increase parameter estimate
  - decrease standard error

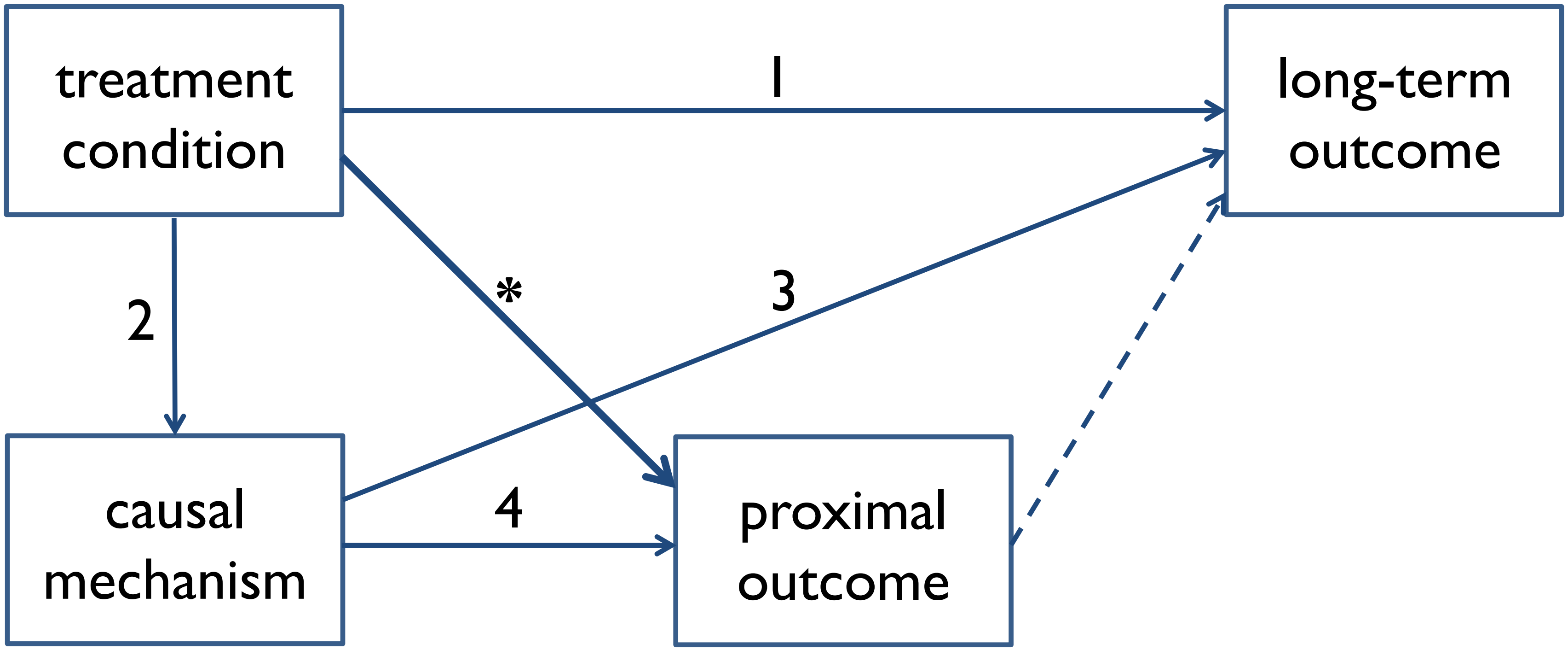
Increasing parameter estimate

$$T = \frac{parameter\ estimate}{standard\ error}$$

- strengthen treatment condition
  - increase "dosage" of treatment
  - choose inactive control condition
- choose reliable but sensitive outcome measure
  - minimize attenuation due to unreliability
  - maximize odds of detecting difference or change by using outcome that is responsive to change in conditions
- focus treatment directly on causal mechanism

Increasing parameter estimate

$$T = \frac{parameter\ estimate}{standard\ error}$$



Decreasing standard error

$$T = \frac{parameter\ estimate}{standard\ error}$$

- examples of standard error

$$\sigma_M = \sqrt{\frac{s^2}{n}}$$

$$\sigma_{M_1 - M_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- options when sample size is constrained . . .

Leave no data unanalyzed

$$T = \frac{parameter\ estimate}{standard\ error}$$

$$\sigma_M = \sqrt{\frac{s^2}{n}}$$

- ensure that the full sample is the analysis sample
  - minimize attrition in prospective studies
  - use modern methods for managing missing data
    - multiple imputation
    - model-based methods
      - e.g., FIML in SEM
      - incorporate missing data mechanism in model
      - inclusion of auxiliary variables

Account for unexplained variance in outcome

$$Y = \beta_0 + \beta_1 x_1 + e_F$$

$$F = \frac{(e_R - e_F)/(df_R - df_F)}{e_F/df_F}$$

- reduce e_F by including covariates associated with e_F (i.e., variance in *Y* not accounted for by predictors of interest)

$$Y = \beta_0 + \beta_1 x_1 + \underbrace{\beta_2 x_{c1} \ldots \beta_i x_{cj}}_{\text{covariates}} + e_F \downarrow$$

Multivariate models

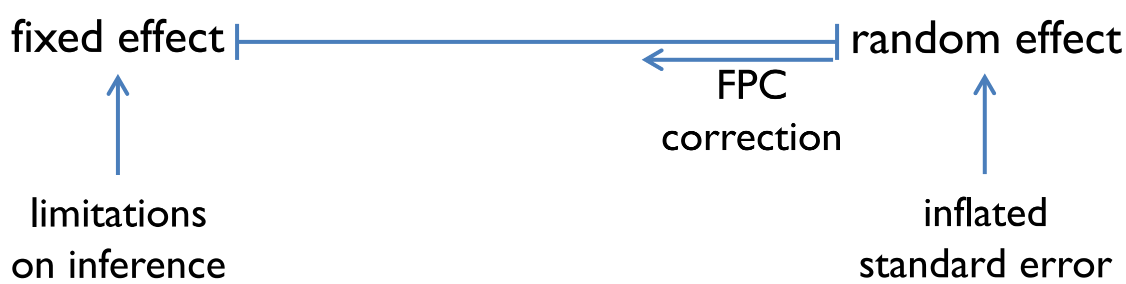
- generally considered large sample methods
- yet, increasing evidence of use with small samples
- reviews of behavioral science applications
  - N of higher-level groups < 30 for 21% of MLM studies
  - $N < 100$ for 33% of growth models
  - $N < 100$ for 40% and < 200 for 60% of EFA studies
  - $N < 100$ for 18% of SEM studies
- suggests research questions that, despite small N, . . .
  - require data that are clustered;
  - concern unobserved influences;
  - focus on patterns of change over repeated assessments

Multilevel modeling

- making the following assumptions
  - continuous measures
  - ICC ≈ .20
  - 4-8 predictors
  - no missing data
  - 2 or fewer cluster-level random effects
- < 40 clusters is considered small
- < 20 clusters should not be analyzed using standard methods
- clusters should have at least 5 observations

Multilevel modeling

- solutions (McNeish, 2017)
  - restricted maximum likelihood (REML)
  - REML with Kenward-Roger correction
  - wild cluster bootstrap
  - application of FPC when population of units responsible for clustering is finite (Lai et al., in press)



Growth modeling

- making the following assumptions
  - continuous measures
  - 4-8 observations per person
  - random intercepts and slopes
  - linear growth
  - < 5 time-varying covariates
- $N < 100$ is considered small
- $N < 50$ should not be analyzed using standard methods

Growth modeling

- solutions
  - depends on analytic framework
  - typically SEM, discussed next
  - move to single-case design framework (Moeyaert et al., 2017)

Table 1
Summary of the reviewed studies.

First author	PY	Children with ASD	Males	Females	Age (mean)	Age (min, max)	Research design(s)	Effect size	Intervention model
Apple, A. L.—exp1	2005	2	2	0	5	5	3+2	0.99	Video-modeling
Apple, A. L.—exp2	2005	2 (3)[a]	1 (2)[a]	1	4.5	4–5	3	0.93	Video-modeling
Chung, K.	2007	4	4	0	6.5	6–7	1	1.16	Peer-mediated
Gonzalez-Lopez, A.	1997	4	2	2	6	5–7	2	0.71	Peer-mediated
Kohler, F. W.	1995	3	3	0	4	4	2	1.48	Peer-mediated
Kohler, F. W.	2007	1	0	1	4	4	3	1.63	Peer-mediated
Laushey, K. M.	2000	2	2	0	5	5	2	1.64	Peer-mediated
Maione, L.	2006	1	1	0	5	5	3	1.22	Video-modeling
Mundschenk, N. A.	1995	3	2	1	8.67	7–10	2	1.47	Peer-mediated
Nikopoulos, C. K.	2003	5 (7)[a]	4 (6)[a]	1	11.2	9–15	1, 2	0.65	Video-modeling
Pierce, K.	1997	2	2	0	7.5	7–8	3	1.41	Peer-mediated
Reeve, S. A.	2007	4	3	1	5.75	5–6	1	2.31	Video-modeling
Strain, P. S.	1995	5	5	0	4.8	4–6	2	1.11	Peer-mediated
Thiemann, K. S.	2004	5	5	0	7	6–9	3	1.09	Peer-mediated

Note. PY = publication year; research design(s): 1. AB; 2. reversal; 3. multiple baseline; 4. alternating.
[a] The number outside the parenthesis is the number of the children included in the meta-analysis, and the number inside the parenthesis indicates the original number of the children in the study.

Structural equation modeling

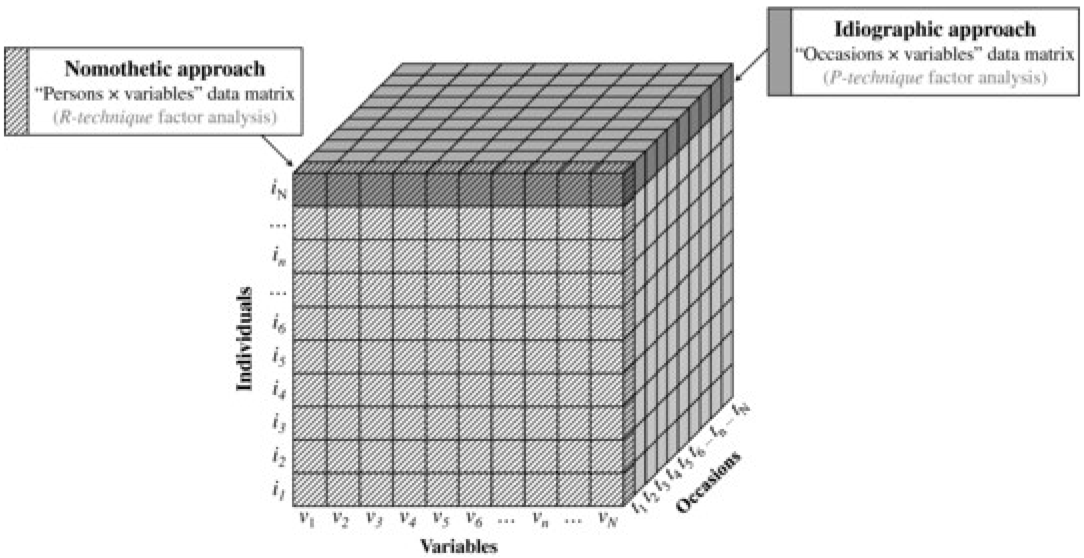
- making the following assumptions
  - continuous measures
  - near-normal multivariate distribution
- and considering the following model characteristics
  - magnitude of loadings on latent variables
  - number of latent variables
  - number of indicators per latent variable
- N < 200 is considered is small for moderate loadings (.5-.7) and moderately complex models (3-4 indicators of 3-4 latent variables)
- N < 100 should not be analyzed using standard methods

Structural equation modeling

- solutions
  - do not interpret raw "χ^2" value; use Bartlett, Yuan, or Swain correction, which include *N*; Yuan-correction performs well at *N*s of 25 and 50
  - use adjunct fit indices that perform well with small *N*s—e.g., Comparative Fit Index
  - limit model complexity

Person-level dynamic modeling

- *p*-technique factor analysis—modeling of intraindividual variability across intensive, repeated observations



Person-level dynamic modeling

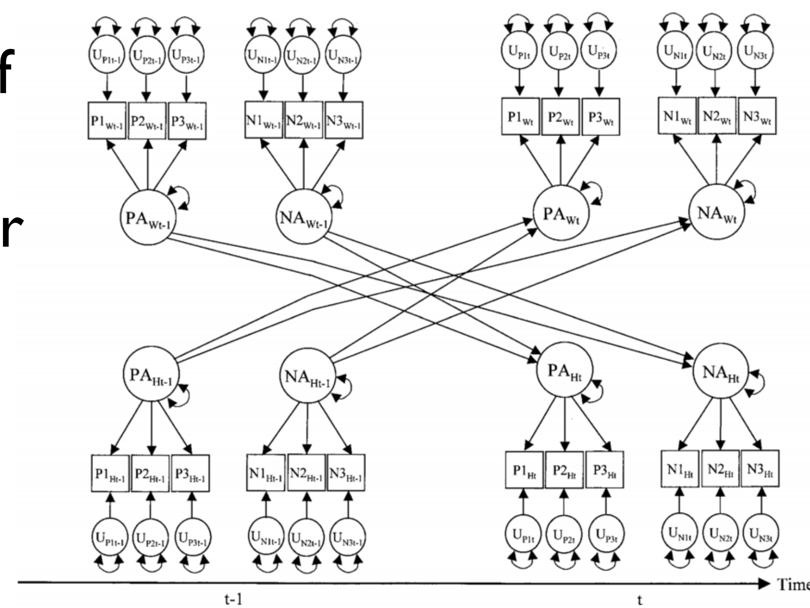
- traditional *p*-technique factor analysis
  - "sample" is a set of observations of one person (*p*) on a set of variables (e.g., measure of affect administered daily for three months yielding ≈ 90 observations)
  - factor analysis of latent structure for the person
  - multiple people can be "chained"
  - limitation: does not account for effects of time

Person-level dynamic modeling

- dynamic *p*-technique factor analysis (Nelson et al., 2011)
  - explicitly incorporates time to allow for modeling of intraindividual change over time
  - uses lagged covariance matrices, permitting modeling of
    - within-lag covariances between variables
    - autoregressive covariances (stability)
    - cross-lagged covariances (prospective relations between variables)
  - person-level data can be chained or analyses done using multigroup SEM

Person-level dynamic modeling

- example: Ferrer & Nesselroade (2003), "Modeling Affective Processes in Dyadic Relations via Dynamic Factor Analysis"
- method
  - one heterosexual couple
  - daily positive & negative affect for 6 months (182 days)
- results
  - different configuration of affect and different trajectories of change for husband and wife
  - unidirectional lagged effect of husband's affect on wife's affect



Person-level dynamic modeling

"DPT can allow for complex models that match the complexity of research hypotheses. Simply stated, DPT allows researchers to conduct sophisticated analyses, despite small numbers of participants. . . . Repeated measurement of a small number of individuals over time is often more feasible than studying large numbers of participants."

(Nelson, Aylward, & Rausch, 2011)

Summary

- the outcome of statistical analysis/modeling should be informative; informative results are challenging to produce for small sample data
- what qualifies as a small sample varies as a function of a number of features of a study
- when *N* is small and constrained, the goal is to maximize the yield of the study through careful consideration of design, measurement, and analysis options
- health research often concerns patterns, processes, or structures that require the use of multivariate methods; such methods can sometimes produce informative results when *N* is small

Thank you!

rhoyle@duke.edu