Statistical Power for Trend Detection Under Alternative Panel Designs for Surveys over Time

Statistical Power for Trend Detection Under Alternative Panel Designs for Surveys over Time

Anthony Olsen, Thomas Kincaid, Steven Paulsen

US EPA ORD NHEERL

Western Ecology Division

Options for Trend Estimation

- Simple and weighted linear regression models
  - Based on annual designed-based estimates of status
- Linear mixed model(s)
  - does not incorporate weights from design (Urquhart et al 1993, Piepho & Ogutu 2002, Urquhart 2012)
- Probability-Weighted Iterative Generalized Least Squares (PWIGLS)
  - Simple fixed- and random effect models applied in design-based setting (Pfeffermann et al. 1998)
- Regional Kendall Tau nonparametric trend detection (Helsel and Frans 2006)
  - Requires all sites be sampled on each sampling occasion. (more general approach by van Belle and Hughes 1984)

Panel/Revisit Designs

- Panel designs are basis for monitoring over time
- Names for panel designs are not consistent in literature but McDonald (2003) provides a naming structure that is helpful

- Always revisit: [1-0] is traditional revisit same sites every time period
- Never revisit: [1-n] is never revisit the same site
- Rotating panel: [3-n] is revisit same site for three years, then never again (rotating panel or sampling with partial replacement)
- Serially alternating: [1-3] is visit site in year one, skip three years and then repeat that pattern in future years
- Split panel: [1-0, 1-4] split panel design with always revisit and serially alternating panels
- Complex rotating panel: [1-3-1-3-1-n] panel design: panel 1 - visit site in years one, five and nine and then never visit again

[1-0] Panel Design: Always Revisit

	YR 1	YR 2	YR 3	YR 4	YR 5	YR 6	YR 7	YR 8	YR 9	YR 10	ETC
Panel 1	150	150	150	150	150	150	150	150	150	150	
Cum#	150	150	150	150	150	150	150	150	150	150	Etc

- **Visit same 150 sites in all years.**
- **No additional site evaluation after first year**
- **Landowners may begin to deny access**
- **Sites may be impacted by sampling**
- **Sites may be "treated" differently by owner or by environmental agency.**

[1-n] Panel Design: Never Revisit

	YR 1	YR 2	YR 3	YR 4	YR 5	YR 6	ETC				
Panel 1	150										
Panel 2		150									
Panel 3			150								
Panel 4				150							
Panel 5					150						
Panel 6						150					
etc											
Cum#	150	300	450	600	750	900	1050	1200	1350	1500	etc

- **New sites every year**
- **Same level of site evaluation effort each year**
- **Landowners only asked for permission for one time**
- **Impact due to sampling or being "treated" differently by owner or by environmental agency not an issue**

[3-n] Panel Design: Rotating Panel

	YR 1	YR 2	YR 3	YR 4	YR 5	YR 6	YR 7	YR 8	ETC		
Panel 1	50	50	50								
Panel 2		50	50	50							
Panel 3			50	50	50						
Panel 4				50	50	50					
Panel 5					50	50	50				
Panel 6						50	50	50			
etc											
Cum#	50	100	150	200	250	300	350	400	450	500	etc

- **Visit sites for three years in row and then never again, i. e., drop 50 sites each year and add new 50 sites.**
- **Number of unique sites increases each year but not as fast as when never revisit a site.**
- **Landowner only has three visits to site which may impact whether they give permission to sample more than once.**
- **Note that if want to always sample 150 sites each year, then will have a set of "start up" panels that add additional sites in first 2 years.**

[1-3] Panel Design: Serially Alternating

	YR 1	YR 2	YR 3	YR 4	YR 5	YR 6	YR 7	YR 8	YR 9	YR 10	YR 11	YR 12	etc
Panel 1	150				150				150				
Panel 2		150				150				150			
Panel 3			150				150				150		
Panel 4				150				150				150	
Cum#	150	300	450	600	600	600	600	600	600	600	600	600	etc

- **Visit 150 sites in year 1, again in year 5 and again in year 9 and continue this serially alternating pattern. Other panels same except start in year 2, 3 and 4.**
- **Reduces issue with landowner access by not sampling every year**
- **Likely reduces impact on site**
- **Potential of "improving" sites sampled if they are in poor condition either by landowners or environment agency**

[1-3-1-3-1-n] Panel Design

	YR 1	YR 2	YR 3	YR 4	YR 5	YR 6	YR 7	YR 8	YR 9	YR 10	YR 11	YR 12	YR 13	YR 14
Panel 1	50				50				50					
Panel 2		50				50				50				
Panel 3			50				50				50			
Panel 4				50				50				50		
Panel 5					50				50				50	
Panel 6						50				50				50
etc														
Cum#	50	100	150	200	250	300	350	400	450	500	550	600	etc	

- **Visit 50 sites in year 1, year 5 and year 9 and never again; etc for other panels**
- **Cumulative number of sites increases over time so that visit more unique sites over time.**
- **Maximum number of times a site is visited is three.**
- **Reduces issue with landowner access, impact on site, and potential of "improving" just sites sampled if they are in poor condition.**
- **Note that if want to always sample 150 sites each year, then will have a set of "start up" panels that add additional sites in first 9 years.**

Considerations: Panel Design Choice

- Power to detect trends in status over time
- Monitoring organization's time/cost to complete site evaluations required each year
- Likelihood landowner will give permission to sample a site when asked to give permission for multiple visits
- Impact sampling has on a site by returning to the same site multiple times
- Potential for landowner to impact site (positively or negatively) when they know site will be sampled multiple times
- Potential for environmental agency to improve a site given that they know its current condition

Power to Detect Trends in Status

Mixed linear model

$$y_{ijk} = \beta_0 + w_j\beta_1 + T_j + S_i + w_jB_i + C_{ij} + \varepsilon_{ijk}$$

$i = 1,\ldots,s$ number of sites
$j = 1,\ldots,t$ number of years
$k = 1,\ldots,r$ number of replications within site and year
w_j - location shifted year variable representing jth year
β_0 and β_1 – fixed intercept and slope of linear trend in time
T_j - random effect of *j*th year
S_i and B_i - random effects of *i*th site intercept and trend slope
C_{ij} - random effect of *i*th site and *j*th year interaction
ε_{ijk} - unexplained error

- y may be Benthic macroinvertebrate MMI, total nitrogen, etc.
- Focus is on determining the power when testing null hypothesis that β_1, the overall linear trend in the population, is equal to 0
- Need estimates for variance components T_j , S_i , B_i , C_{ij} and ε_{ijk}
- Can simplify model: drop B_i (variation adsorbed into C_{ij}); drop ε_{ijk} if no replications

Variance Components Required

$$\sigma^2_{Tot} = \sigma^2_{Site} + \sigma^2_{Year} + \sigma^2_{Site*Year} + \sigma^2_{SiteTrend} + \sigma^2_{Error}$$

$\hat{\beta} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}Y$ and $cov(\hat{\beta}) = (X'\Sigma^{-1}X)^{-1}$

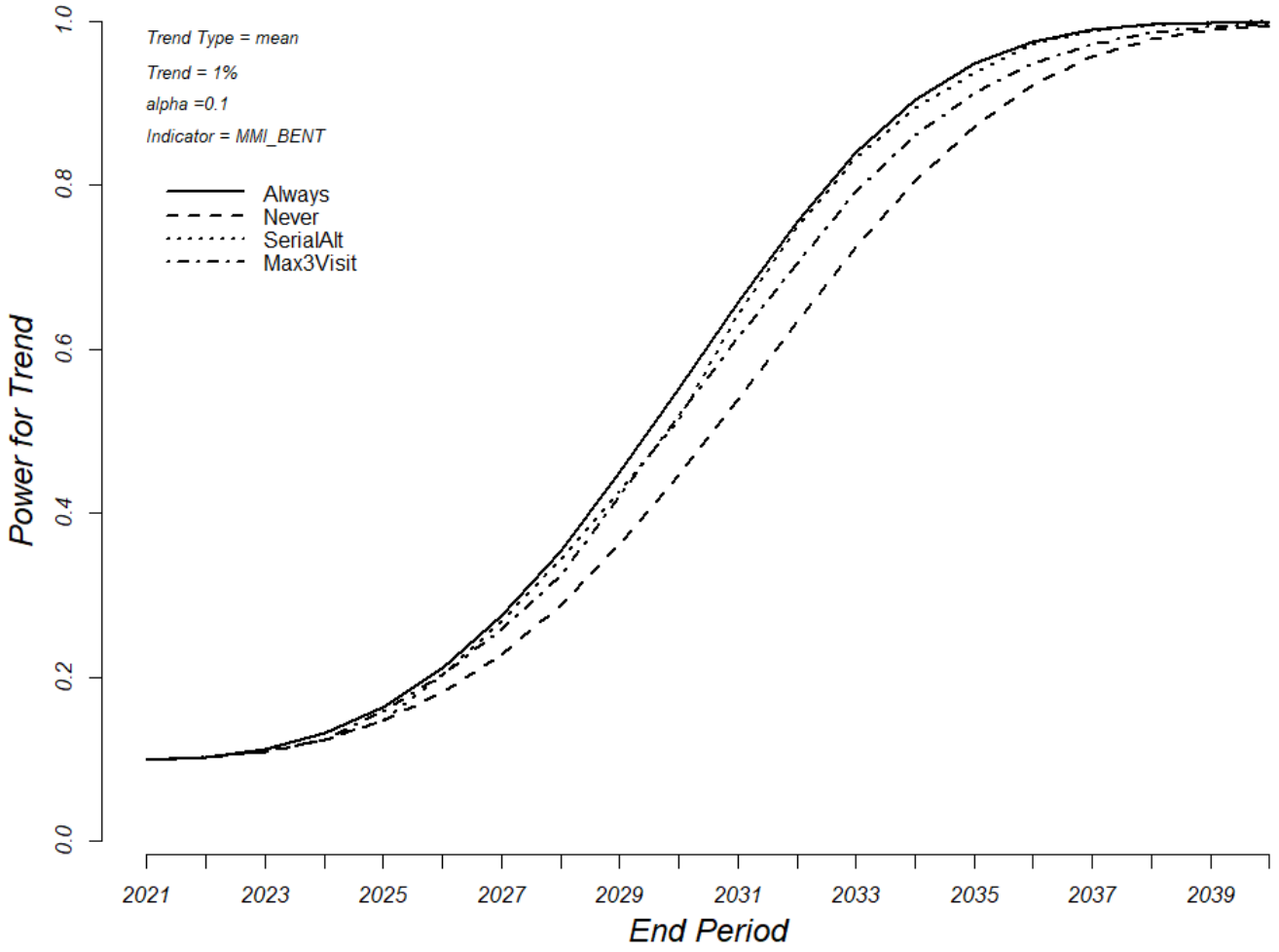
X is matrix with column of ones and column containing time period variable for year, w_j, and

Σ is cov(Y) where Y is vector of responses Y_{ij} and includes information on panel designs and variance components

$\hat{\beta}_1$, second element of $\hat{\beta}$, is estimated population trend and its standard error is the square root of the second diagonal element of $cov(\hat{\beta})$, β_1^0 is assumed true trend, Φ is cumulative normal distribution, and power is expected power given true trend, panel design and sample size.

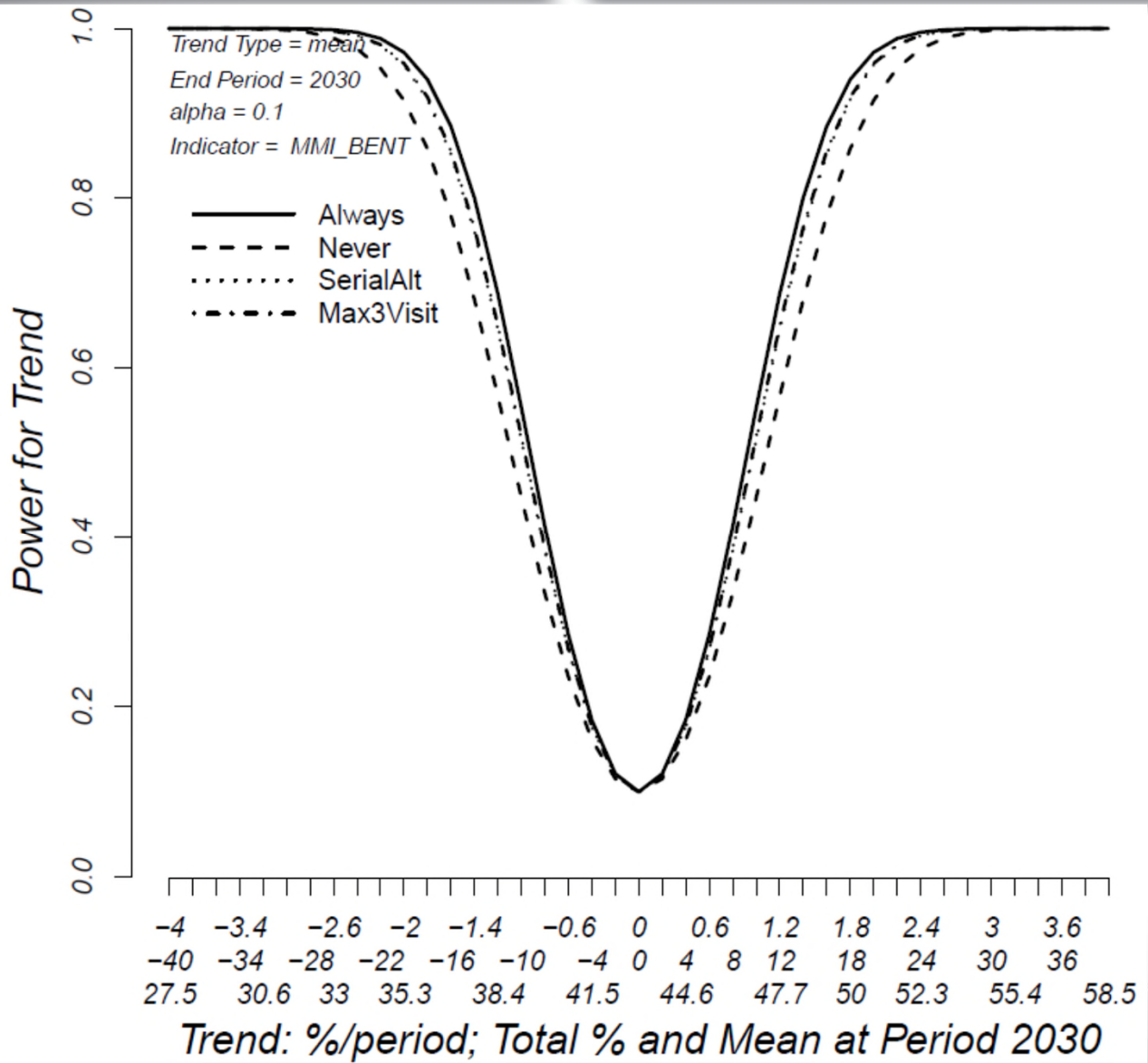
$$\text{Power}(\beta_1^0) = \Phi\left(-1.645 - \frac{\beta_1^0}{\sigma_{\hat{\beta}}}\right) + 1 - \Phi\left(1.645 - \frac{\beta_1^0}{\sigma_{\hat{\beta}}}\right)$$

Power Four Panel Designs



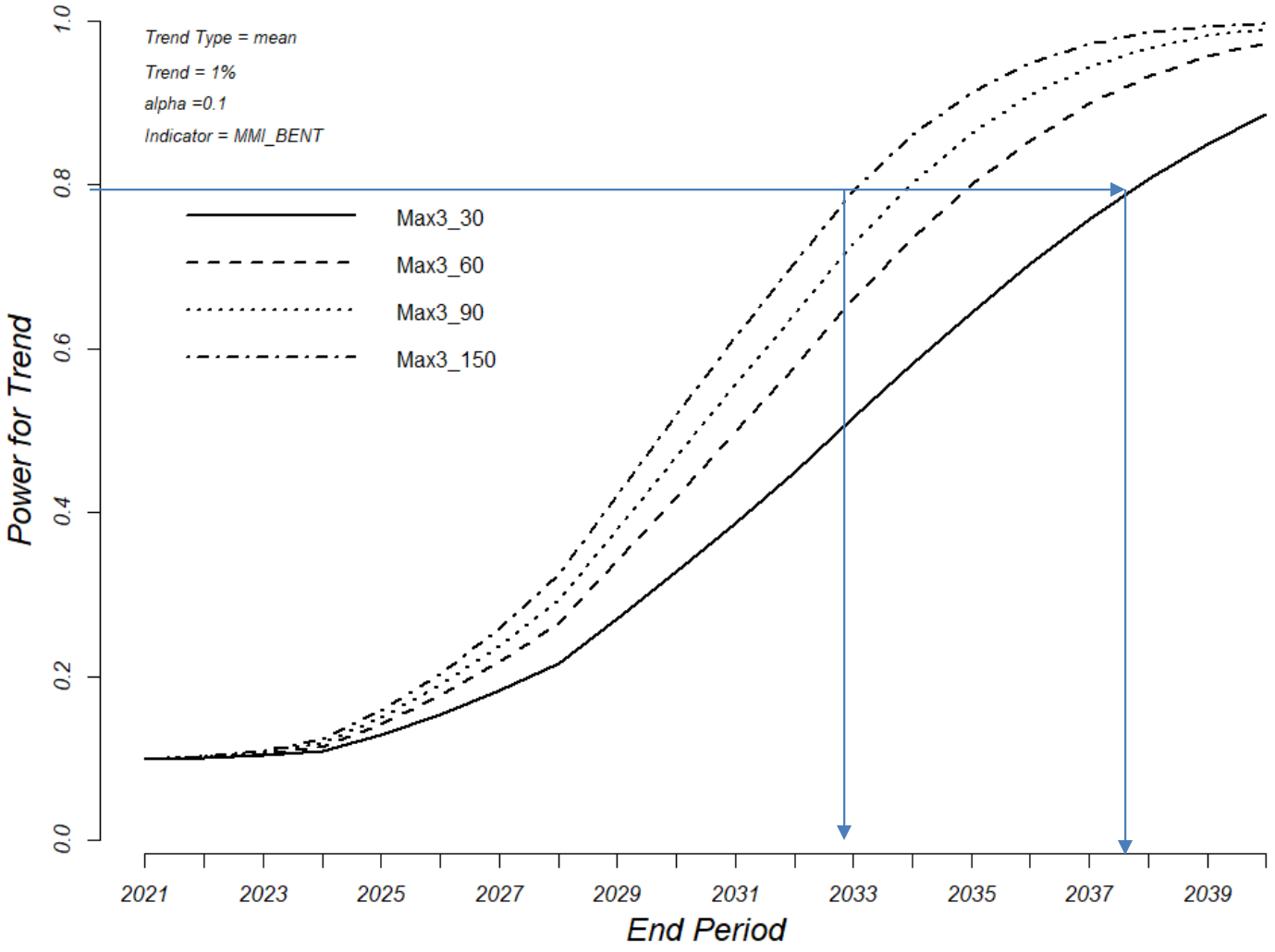
- **Four panel designs**
  - **Always [1-0]**
  - **Never [1-n]**
  - **SerialAlt [1-4]**
  - **Max3Visit [1-3-1-3-1-n]**
- **Sample size 150 with Trend 1%/Year**
- **Power by year**
- **Power is similar for all panel designs but Never revisit panel design has lowest power**

Power Four Panel Designs



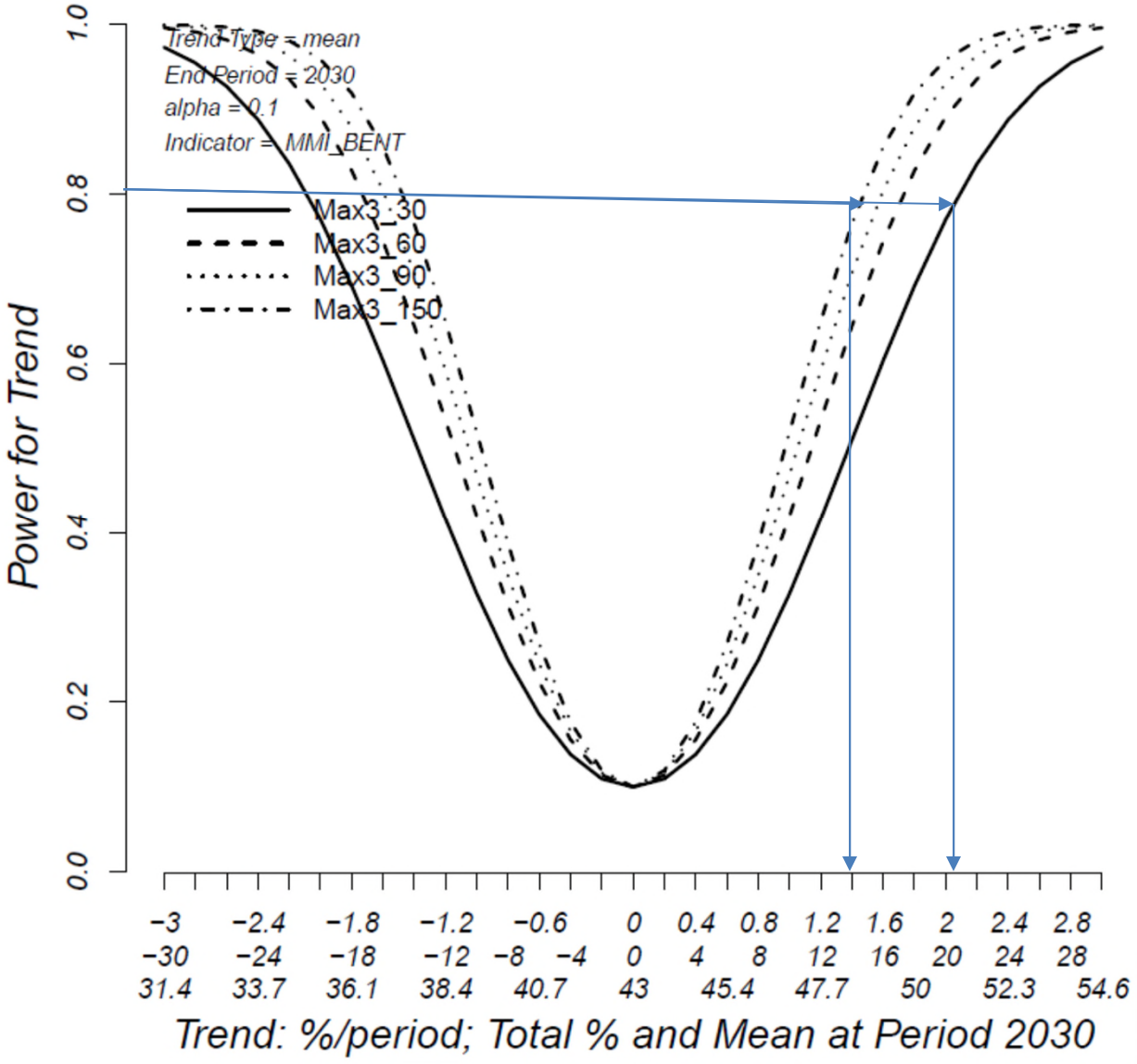
- **Four panel designs**
  - **Always [1-0]**
  - **Never [1-n]**
  - **SerialAlt [1-4]**
  - **Max3Visit [1-3-1-3-1-n]**
- **Sample size 150**
- **Power after 10 years**
- **Power is similar for all panel designs but Never revisit panel design has lowest power**
- **Panel design choice may depend on other factors**

[1-3-1-3-1-n] Power Sample Size



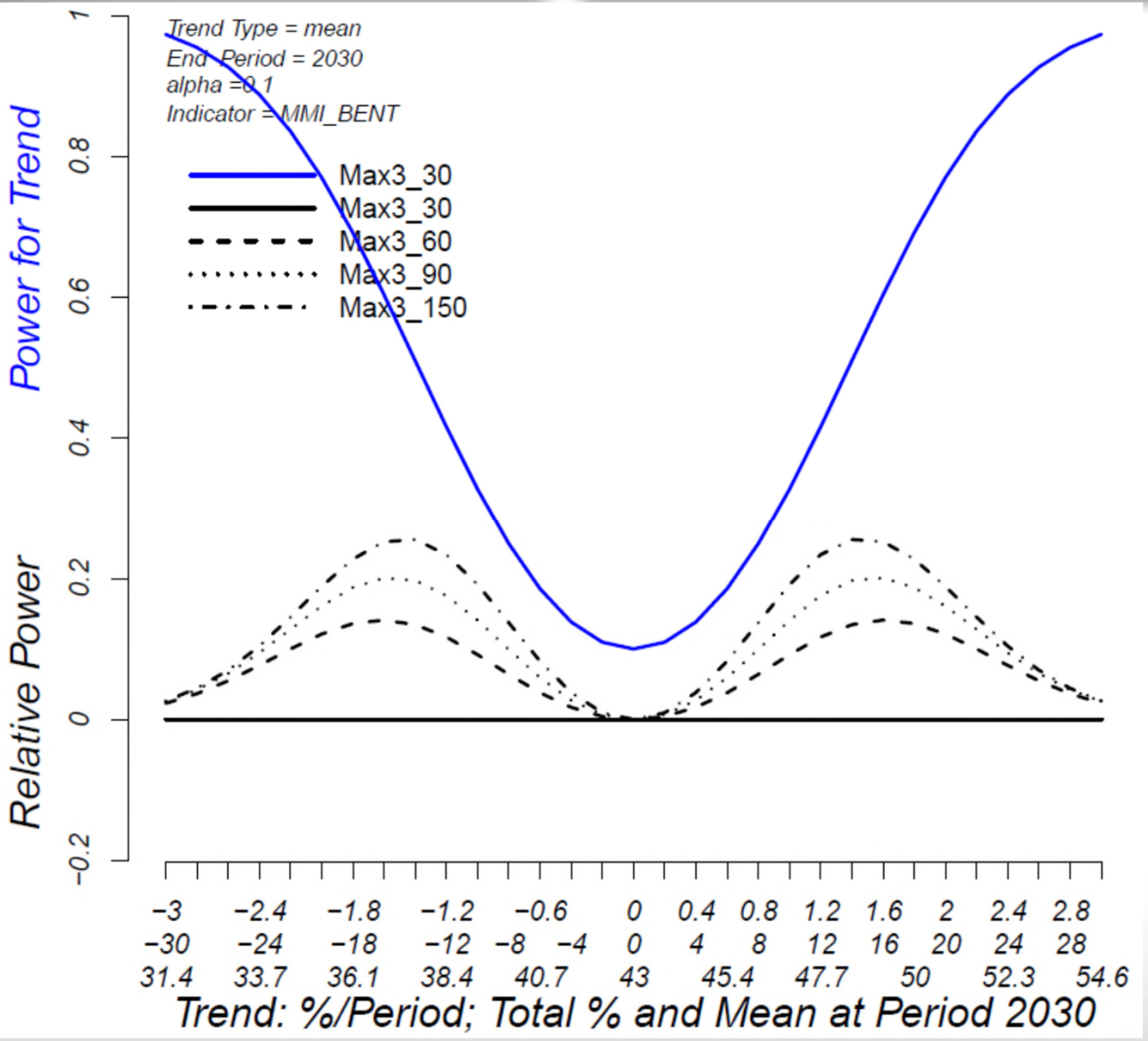
- **Power by years monitored for sample sizes of 30, 60, 90 and 150**
- **Trend: 1%/year. After 10 years a total change of 10%**
- **Power > 0.8 in 2033 if n = 150 and Power > 0.8 in 2038 if n = 30**

[1-3-1-3-1-n] Power Sample Size: 10 Years



- **After 10 years (2030), power is >0.8 for all sample sizes if trend is >2%/year (or <-2%/year). At 10 years total % change is 20% (or -20%) and MMI changes to ~51 (or ~35) from base of 43.**
- **If sample size is 150/year, then power >0.8 if trend is >1.4% or total % change 14% with MMI change to ~49 from base of 43.**

[1-3-1-3-1-n] Relative Power Sample Size



- **If use panel design with 30 sites as base, then increase in power if increase sample size to 60, 90 and 150 is shown in black.**
- **If increase sample size by factor of five (to 150), then largest power increase is ~0.2 when %/year is ~1.5%**
- **Can see on prior plot but this makes explicit comparison.**

Power Computation Alternatives

- **Power Based on Variance Components**
  - **Alternative panel design alternatives included**
  - **Assumes survey design is simple random sample, i.e., complex survey design structure not accounted for in computation**
  - **Requires estimating variance components based on existing data**
  - **Assumes estimated variance components are known for power analysis**
  - **Power calculations simple to compute given R functions**
- **Power Based on Simulations**
  - **Alternative panel design alternatives included**
  - **Complex survey design structure included in simulation**
  - **Requires creating response values for entire sampling frame under hypothesis of no trend as well as set of alternative trends. Created response values may be based on a model.**
  - **Requires selecting samples large number of times (~1,000) and estimating trend for each**
  - **Power is simply proportion of simulations where null hypothesis of no significant trend is rejected**

R functions for Power

- **Revisit_dsgn: Create panel designs**
  - **Creates a revisit design for panels in a survey that specifies the time periods that members of each panel will be sampled using panel naming convention of McDonald (2003).**
  - **Three basic revisit design structures may be created: always revisit panel, serially alternating panels, or rotating panels**
  - **revisit_bibd and revisit_rand are additional alternatives**
  - **panel_summary: summarizes characteristics of a revisit design**
- **power.dsgn: Compute power for trend detection**
  - **Based on Urquhart (2012) mixed linear model**
  - **power for trend detection for one or more variables, for one or more panel designs, for one or more linear trends, and for one or more significance levels**
- **plot.powerpaneldesign: Plot power based on results from power.dsgn**
  - **Multiple options for creating power plots**
- **All soon to be added to spsurvey**