A Systematic Approach to Search for Efficient Designs in Analysis of Change

Fan Jia, Wei Wu, Mijke Rhemtulla, Todd Little

Modern Modeling Methods, May 22 2014

Outline

- Background
  - Linear and quadratic growth models
  - Two types of designs
  - Literature review
- A systematic approach to search for efficient designs
- Simulation
- Results
- Discussion and future extensions

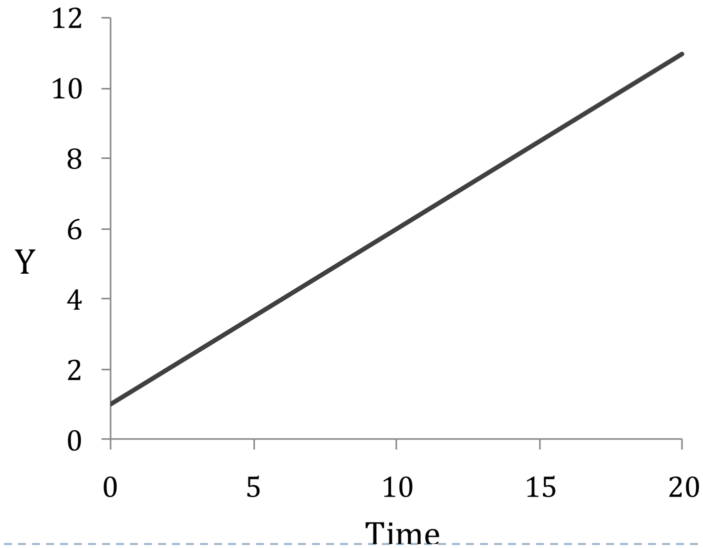
Analysis of Change

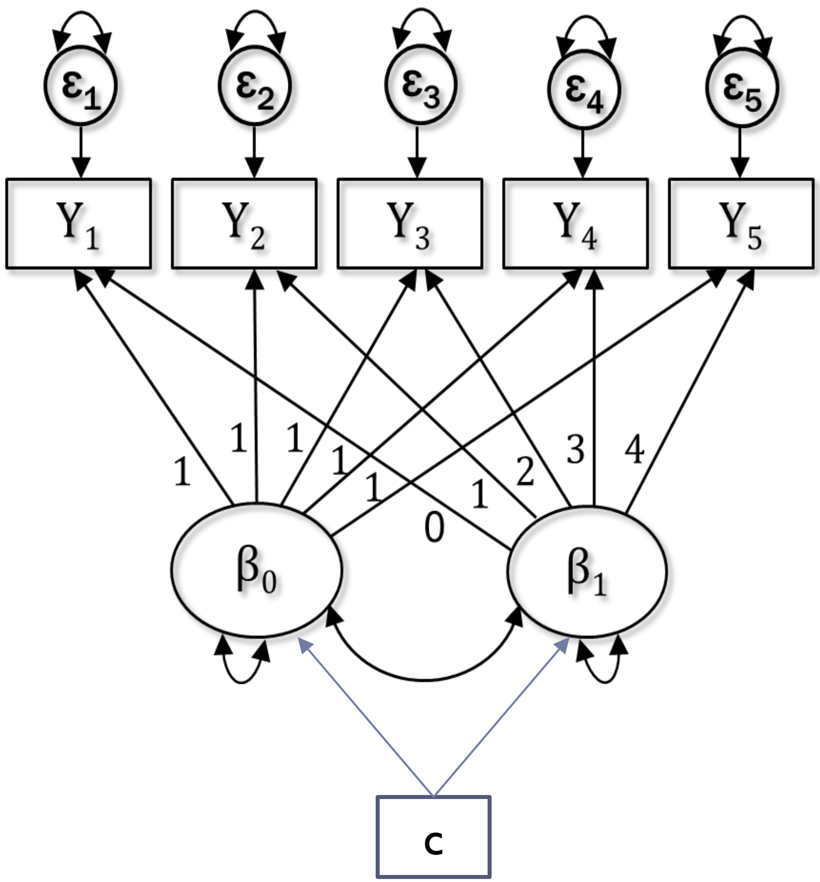
- Change is of central interest to many research areas.
- Require repeated measures.
- Mixed models or SEMs are often used to examine the intra- and inter-individual changes over time.

Linear Growth Curve Model

Fixed effects: mean intercept and slope, covariate effect on intercept and slope

Random effects: variances and covariance of the intercept and linear slope.

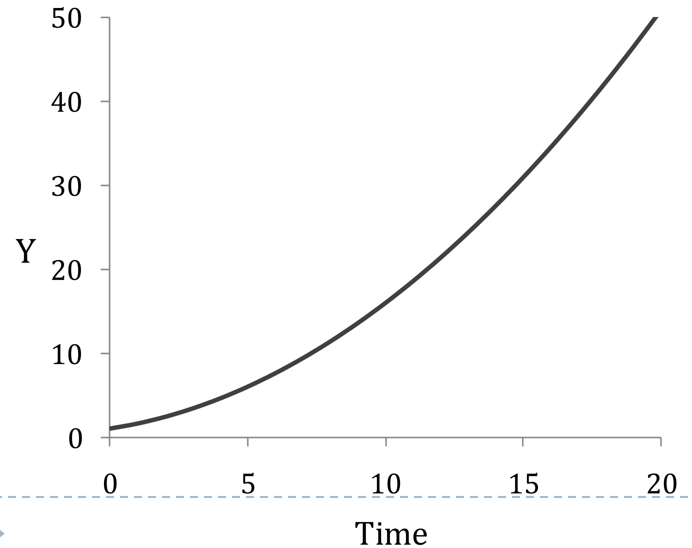


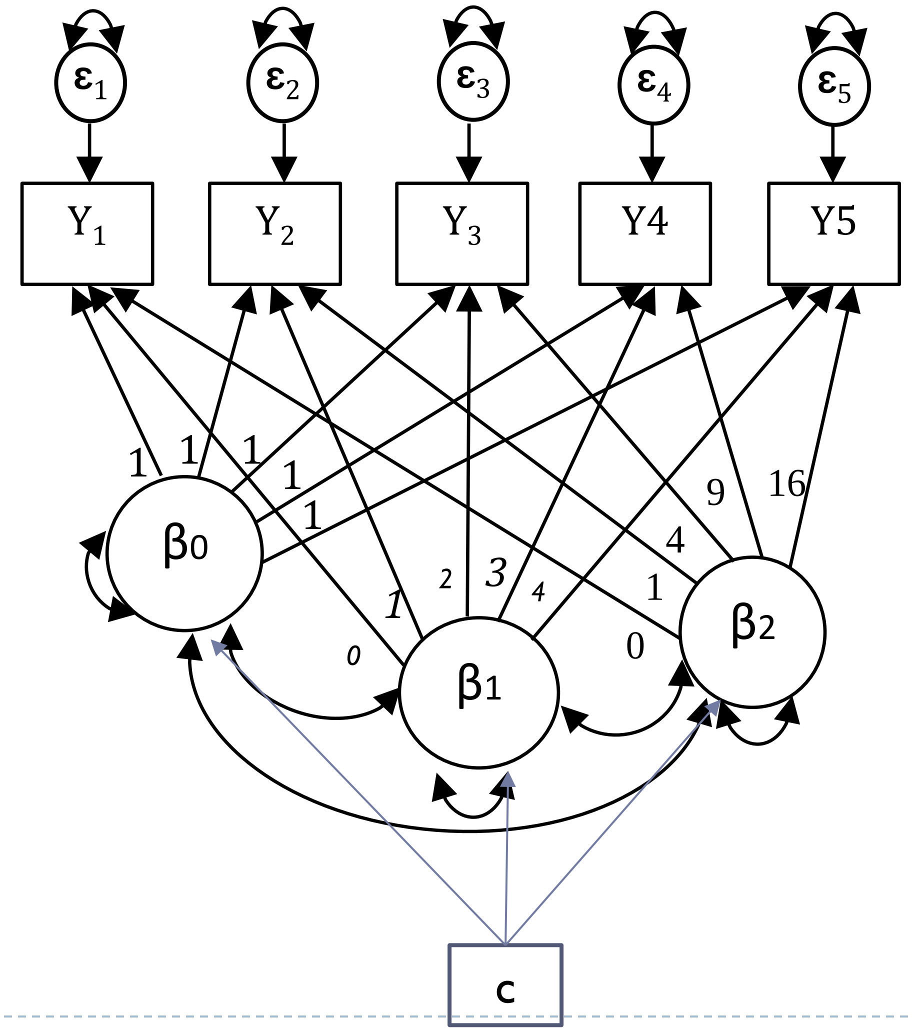


Quadratic Model

Fixed effects: mean intercept, linear slope, and quadratic slope, covariate effects on the intercept and slopes

Random effects: variances and covariances of the intercept, linear slope, and quadratic slope





Two Types of Designs

- **Complete Data Designs**
  - One response pattern
  - Complete data method is used to analyze the data.

- **Wave Level Planned Missing Data Designs (WLPMDs)**
  - Multiple response patterns.
  - Missing data techniques have to be used to analyze the data.

{1,2,3,4,5}

P	Time				
	1	2	3	4	5
1	X	X	X	X	X

{1,2,4,5}

P	Time				
	1	2	3	4	5
1	X	X		X	X

{1,2,3; 2,3,4; 3,4,5}

P	Time				
	1	2	3	4	5
1	X	X	X		
2		X	X	X	
3			X	X	X

{1,4,5; 2,3,5; 1,3,5}

P	Time				
	1	2	3	4	5
1	X			X	X
2		X	X		X
3	X		X		X

Past Research on Complete Data Designs

- In biomedical research, numerous studies have been conducted to identify optimal complete data designs to model change given fixed budget (Abt et al., 1997;1998; Ouwens, Tan, & Berger, 2002 ; Bischoff ;1993; Mentre, Mallet, & Baccar,1997; Tekle, Tan, & Berger; 2011) .

  - Focus on multiple fixed effects in polynomial model.
  - Identify an optimal design using optimization algorithm.
  - Different criteria have been utilized to identify an optimal design.
    - One of the criterion is to minimize the generalized variance of fixed effects.
  - Consider additional attrition rates.

General Findings

- The optimal complete data designs tend to have observations at the first and last measurement occasions.
- Equally spaced time points generally work pretty well.
- Too many repeated measures will result in loss of efficiency due to correlations among repeated measures.

The best number of time points

  - = the number of fixed effects if the tested model fits the data .
  - = the number of fixed effects plus 1 or 2 if the model might not fit the data. (Ortega-Azurdy, Tan, & Berger, 2009; Willet, Singer, & Martin, 1998)

Past Research on WLPMDs

- In social and behavioral science, research has been performed to evaluate the efficiency of WLPMDs in detecting significant linear or nonlinear change which are usually modeled using mixed or multilevel models (Graham, Taylor, & Cumsille, 2001; Mistler, & Enders, 2012).

  - Compare a set of WLPMDs to a specific complete data design.
  - Focus on a single effect at a time.
  - Use empirical approach such as **Monte Carlo** approach to evaluate the efficiency of a design.
  - Conclude that WLPMDs are more cost effective than the complete data design.

Graham et al. (2001)

Graham, Taylor, & Cumsille (2001) examined the efficiency of 7 WLPMDs for a longitudinal study with 5 measurement occasions.

The best design to detect a covariate effect on the slope in a liner model:

Pattern	Measurement occasion						Percentage of sample
	1	2	3	4	5		
1	X	X	X	X	X		9.1%
2	X	X	X				10.1%
3	X	X		X			10.1%
4	X		X	X			10.1%
5	X	X			X		20.2%
6	X		X		X		20.2%
7	X			X	X		20.2%

Mistler and Enders (2012)

Mistler and Enders (2012) compared two WLPMDs to a complete data design with 6 time points in detecting mean linear and quadratic slopes.

The best design:

Pattern	Occasion of measurement							Percentage of Sample
	1	2	3	4	5	6		
1	X	X	X			X		16.7%
2	X	X		X		X		16.7%
3	X		X	X		X		16.7%
4	X	X			X	X		16.7%
5	X		X		X	X		16.7%
6	X			X	X	X		16.7%

More Thoughts on the WLPMD Studies

- A systematic way to search for efficient WLPMDs might be needed.
- Could compare WLPMDs to the optimal complete data designs (rather than the less optimal ones).
- Would additional attrition rate matters?

Purpose of this Study

- Propose a systematic approach to search for efficient designs among a pool of possible complete data and missing data designs **given fixed budget**.
  - utilizing the Monte Carlo method to evaluate the efficiency of a design.
  - taking into account additional attrition.
- This approach is then used to search for efficient designs for a study with 5 measurement occasions.

A Systematic Approach

- **Step 1:** Define the number of measurement occasions.
- **Step 2:** Define the pool of designs.
- **Step 3:** Use Monte Carlo approach to evacuate the efficiency of designs in the design pool for each parameter in the hypothesized model.
- **Step 4:** Summarize the result from step 3.

Step 1

- **Define the total number of measurement occasions (NM).**
  - Minimum unit for time, e.g., month, year
  - Length of study, e.g., 2 or 5 years
  - E.g., unit = half a year, length of study = 2 years, then NM = 5 measurement occasions

Step 2

- **Define the design pool.**
  - To reduce the design pool to a manageable level, it is reasonable to add restrictions to the designs.
  - Complete data designs: 1) observations at the first and last measurement occasions; 2) a minimum number of repeated measures.
  - For missing data designs: 1) participants are equally assigned across missing data patterns ; 2) the response patterns in each design have the same number of missing observations.

Step 3

- **Use Monte Carlo approach to evaluate the efficiency of each design.**
  - Generate a large number of samples (typically 1000) with a specific sample size N from the hypothesized model.
    - Given fixed budget, N varies across designs which can be computed by a function of the total budget, and the number of repeated measures in each response pattern.
  - Impose the planned missingness and additional attrition (if there is any) on the simulated data.

- Fit the simulated data to the analysis model and compute the percent relative efficiency (RE)

$$RE = 100\% \times \frac{VAR(\theta)_{design0}}{VAR(\theta)_{design1}}$$

  - θ indicates a specific parameter.
  - **Var(θ)** is the squared empirical standard error for θ.
  - Design0 is the most efficient design with smallest Var(θ).
  - Design1 is a design which is less efficient than design0.

Step 4

- **Summarize the result.**
  - What are the most efficient designs for the important parameters in linear and quadratic growth curve models?
  - What are the designs that are almost as efficient as the most efficient designs? (designs with RE $\geq$ 90% are treated as efficient designs)
  - What are the designs that are efficient across multiple effects (e.g., all fixed effects for slope factors)?
  - Are the efficient designs robust to additional attrition rates?

Simulation

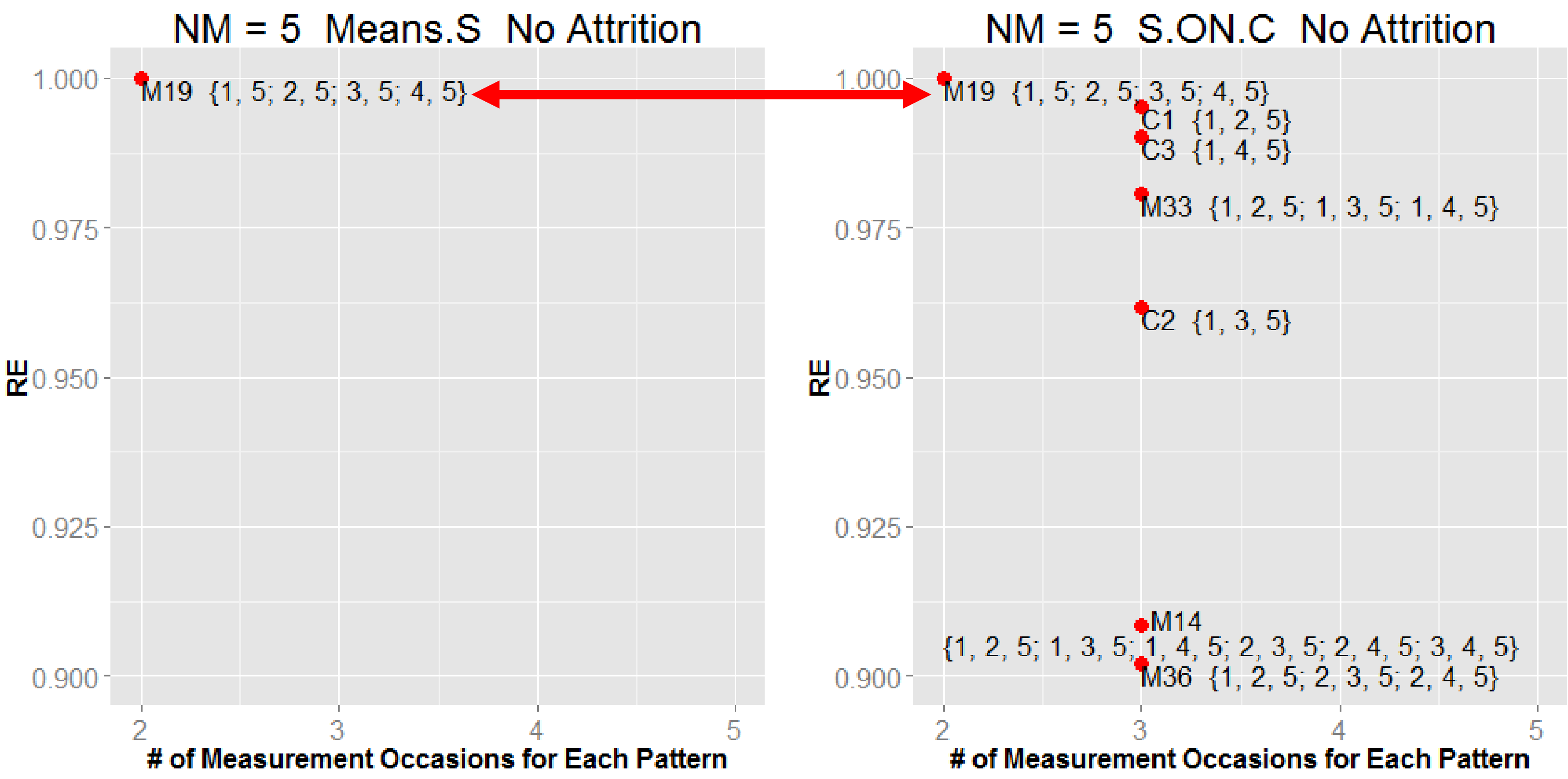
- **NM = 5**
  - The design pool consists of 7 complete data designs and 49 missing data designs.
  - For ease of reference, we assign a label to each design.
  - Example:
    - C3 indicates the third complete data design.
    - M14 indicates the 14th missing data designs.

- **Population models:**
  - Linear and quadratic models with a covariate which has a correlation of .3 with the slope parameter(s).
  - Population values of the parameters are adopted from Biesanz, et al. (2004) and Singer and Willet (2003).

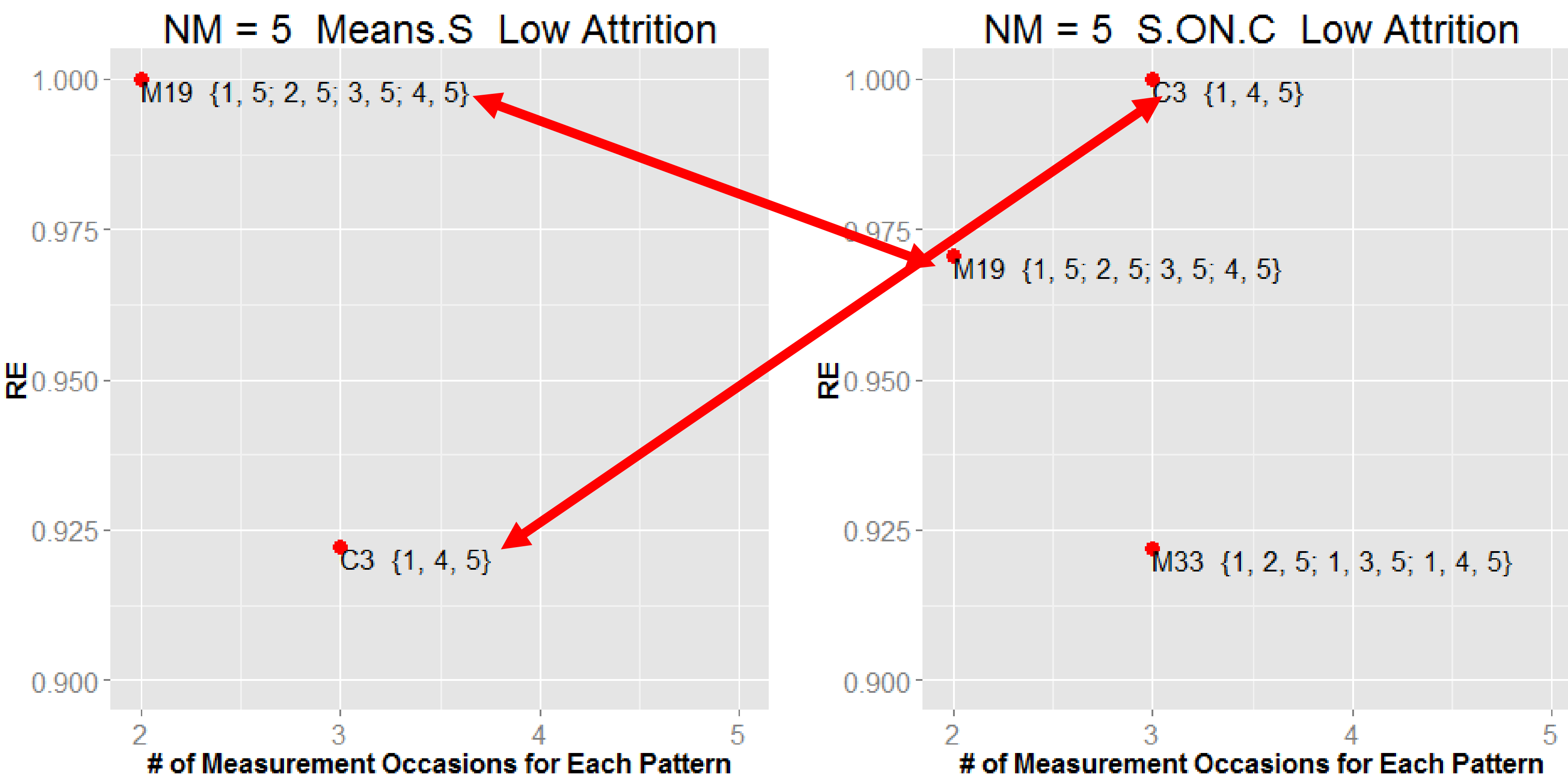
Simulation

- **Simulation factor: level of additional attrition**
  - Attrition rate is a linear function of time (Ortega-Azurduy, Tan, & Berger, 2008).
    - None
    - Low: attrition rates were 0%, 8%, 15%, 23%, and 30%.
    - High: attrition rates were 0%, 18%, 35%, 53%, and 70%.

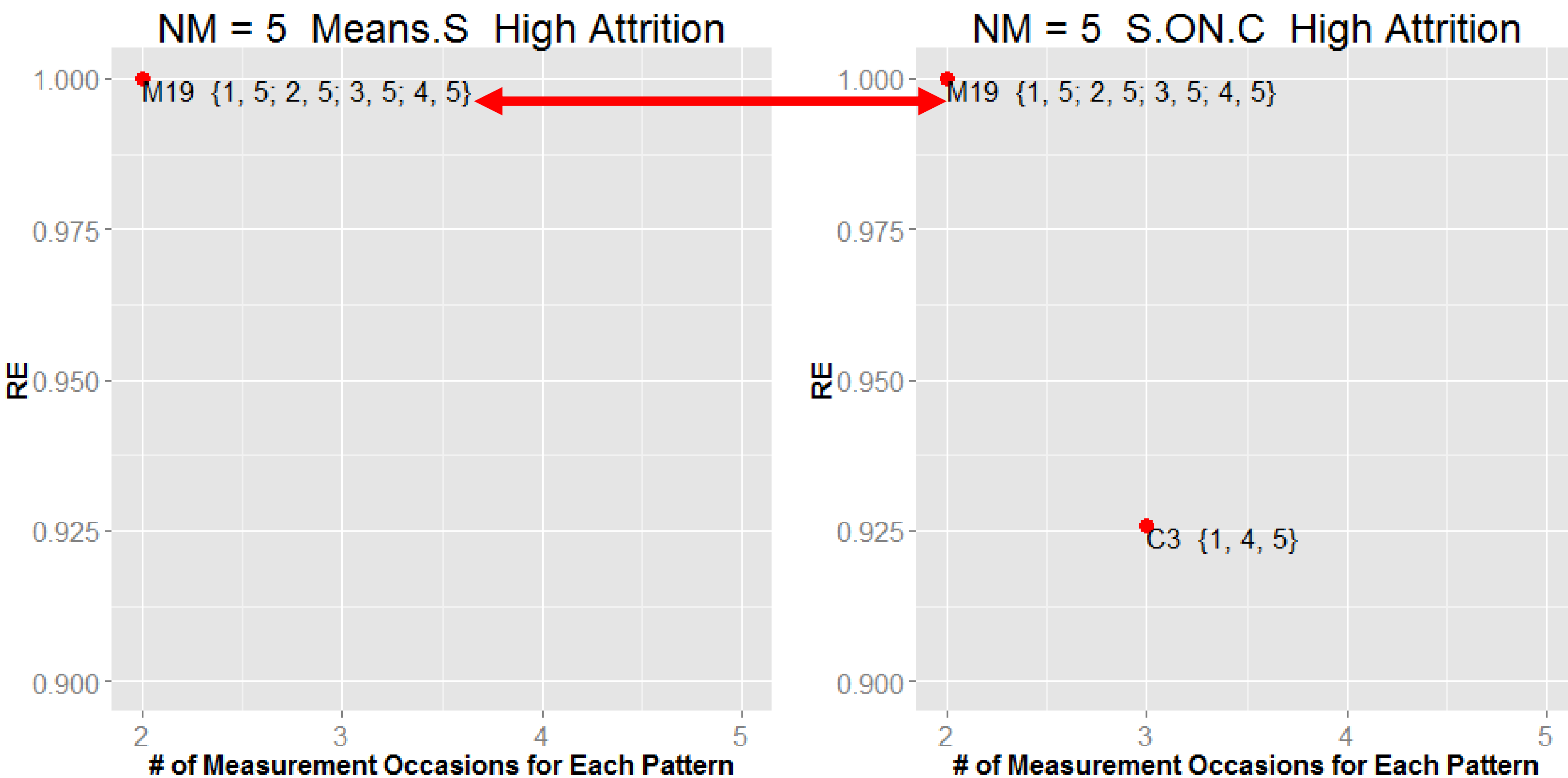
Linear: Efficient Designs without Attrition



Linear: Efficient Designs with Low Attrition



Linear: Efficient Designs with High Attrition



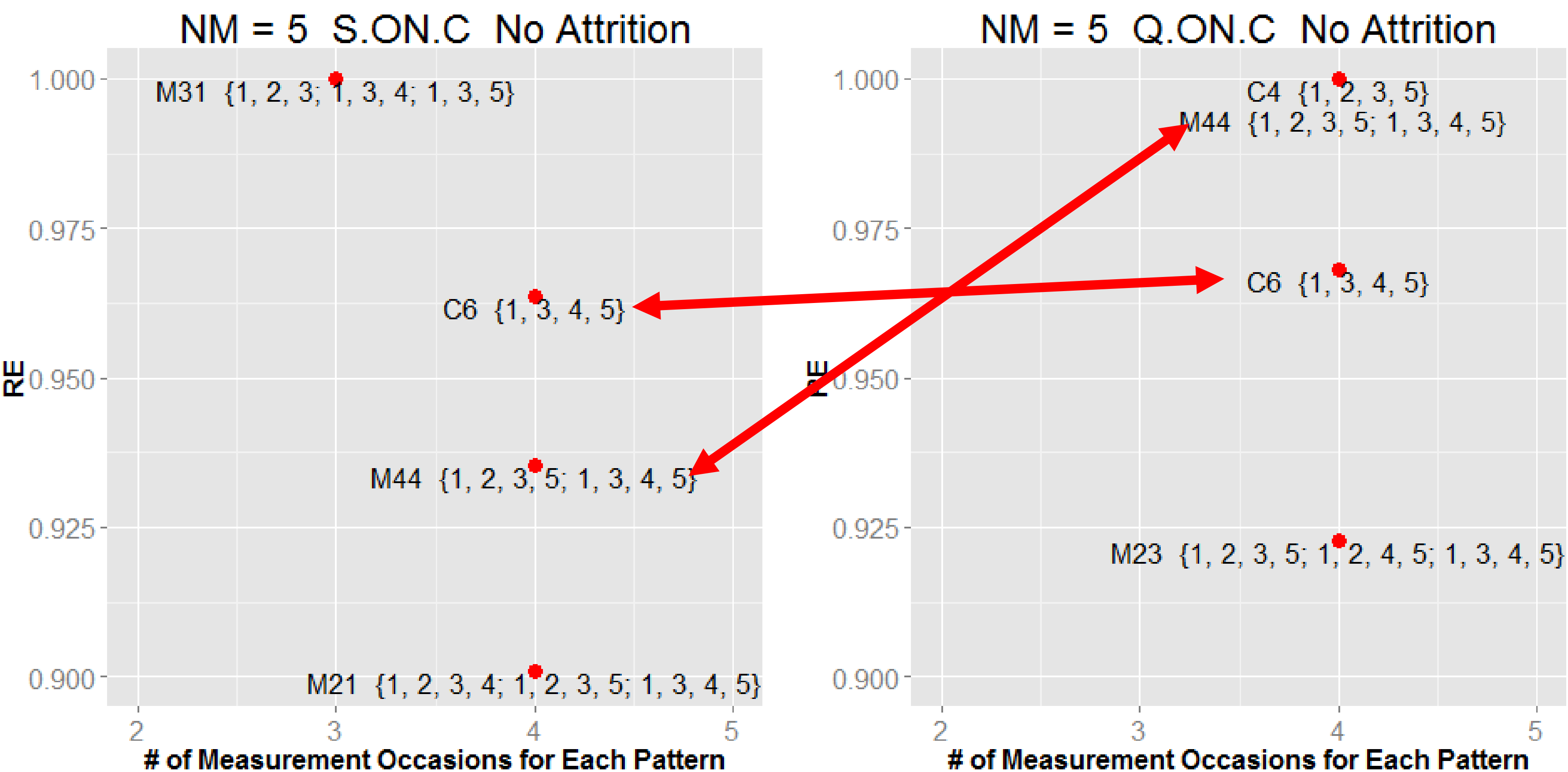
Summary on Linear Model Results

- For the mean slope and covariate effect on slope:

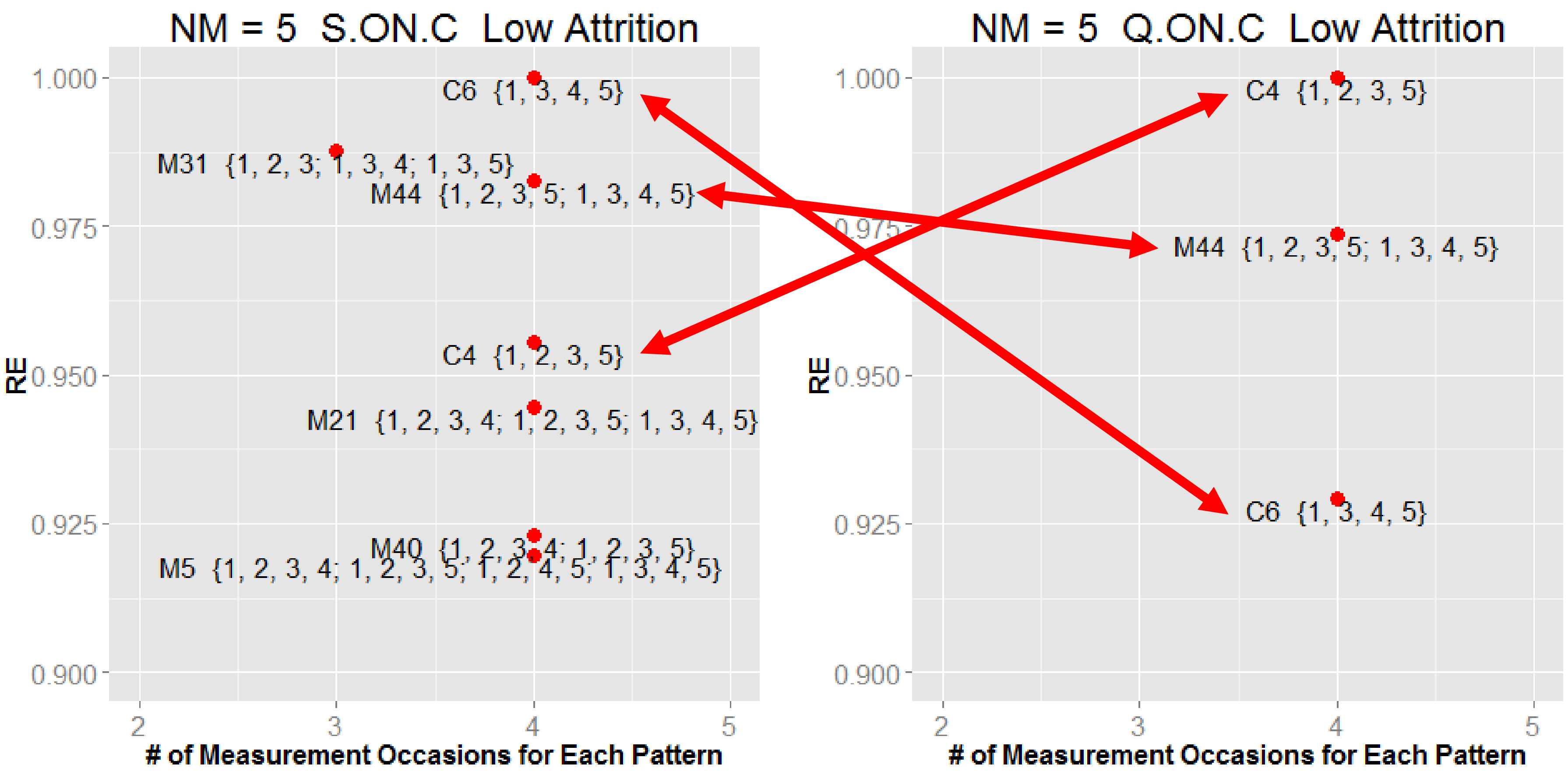
  M19={1,5 ; 2,5; 3,5 ; 4,5} was the most efficient design (with only one exception) , and was robust to change in additional attrition rate.

- There was more than one design that was efficient for a specific parameter based on the threshold of RE = 90% . For example:
  - **C3 ={1, 4, 5}** performed well for estimation of the mean slope with low attrition and for estimation of the covariate effect on slope across different attrition rates.
  - **M33= {1,2,5 ; 1,3,5 ; 1,4,5}** turned out to be an efficient design for estimation of the covariate effect on slope when there was no or low attrition rate.

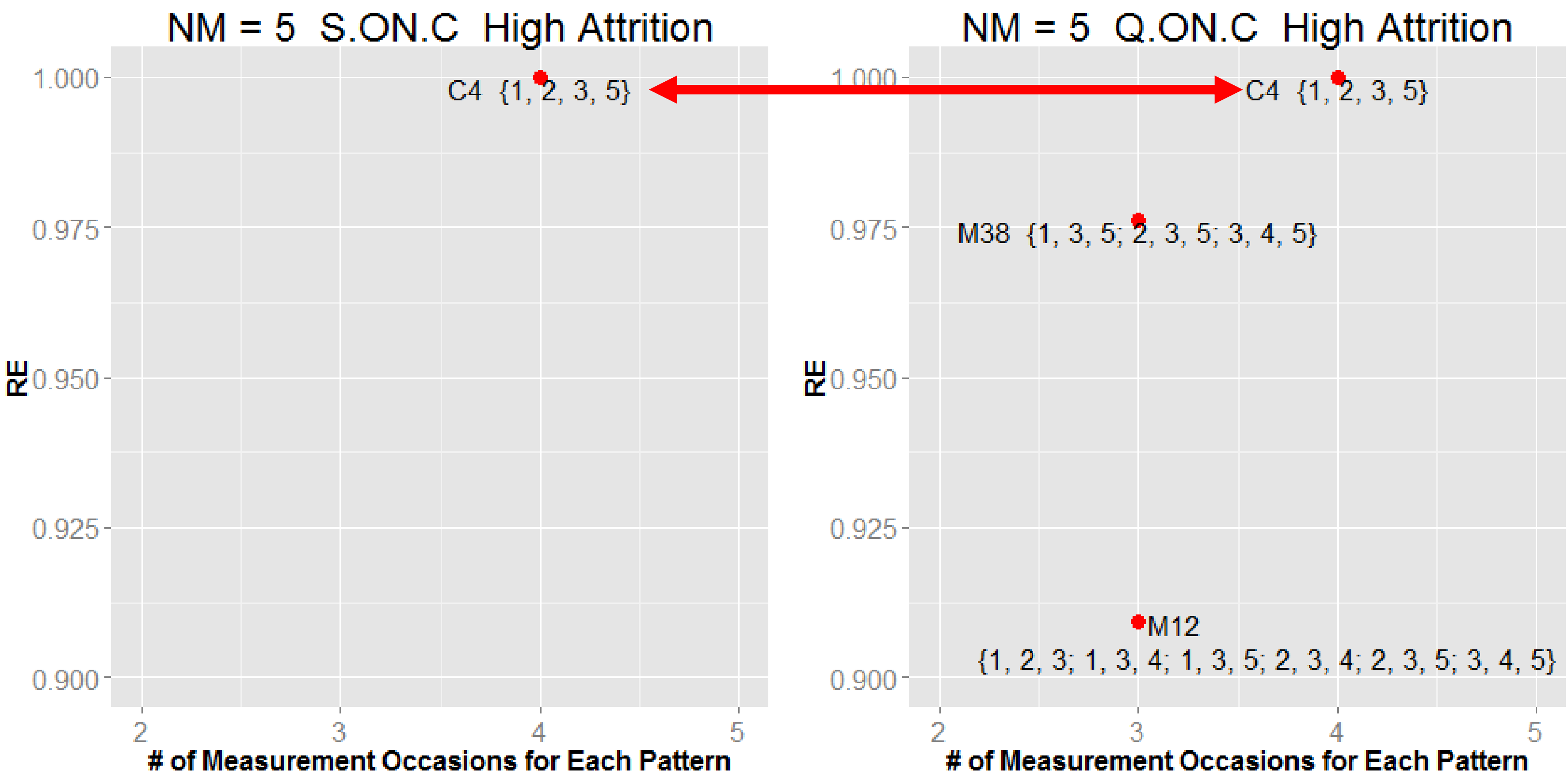
Quadratic: Efficient Designs without Attrition



Quadratic: Efficient Designs with Low Attrition



Quadratic: Efficient Designs with High Attrition



Summary on Quadratic Model Results

- For the mean linear slope and covariate effect on linear slope:
  - The best design varied across different level of attrition rate. **M31={1,2,3 ; 1,3,4; 1,3,5} , C6={1,3,4,5} and C4={1,2,3,5}** turned out to be the best design when there is no attrition, low attrition and high attrition, respectively.
- For the mean quadratic slope and covariate effect on quadratic slope:
  - **C4={1,2,3,5}** was the most efficient design, and was robust to change in additional attrition rate.
- Efficient designs across multiple effects:
  - **M44={1,2,3,5 ; 1,3,4,5} , C6={1,3,4,5}** were the efficient designs across effects when there was no or low additional attrition rate; **C4={1,2,3,5}** performed well for multiple effects when there was low or high additional attrition rate .

The Best Design in Graham et al., (2001)

- For fixed effects in linear model, RE ranged from 72% to 88%.

- For fixed effects in quadratic model, RE ranged from 70% to 90%.

Discussion

- We have proposed a systematic approach to search for efficient designs.
- We have identified some efficient designs to detect fixed effects related to slope(s) in linear and quadratic models with 5 NMs.
- This approach can be easily adopted to different parameter values, different NMs, and different change trajectories.

Future Extensions

- What if the attrition is dependent on previous repeated measures?
- What if the change trajectory is spline?
- The design pool can be further reduced based on the current findings.
- Develop a software package in R or SAS.

Thank You!