Designing Longitudinal Studies with Repeated Measures: the Case of Salivary Cortisol in the Multi-Ethnic Study of Atherosclerosis

Meihua Wu, Brisa N. Sánchez, Trivellore E. Raghunathan
Department of Biostatistics University of Michigan
Ana V. Diez-Roux
Department of Epidemiology University of Michigan

APHA Conference 10/31/2011

Presenter Disclosures

Presenters: Meihua Wu, Brisa N. Sánchez, Trivellore E. Raghunathan, Ana V. Diez-Roux

1. The following personal financial relationships with commercial interests relevant to this presentation existed during the past 12 months:

   No relationships to disclose.

2. This research is supported by Nation Institutes of Health Grant R21 DA024273.

MESA Stress & Salivary Cortisol

- MESA Stress is a large scale epidemiology study which explores the association between stress and cardiovascular disease.
- Incoporate salivary cortisol as an objective and field-friendly measure for stress.

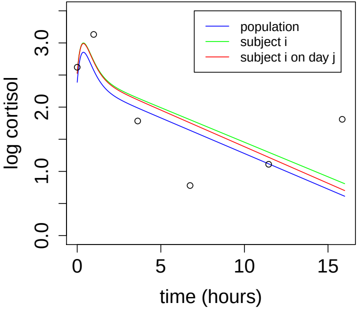


Figure: Cortisol Profile

- Multi-level variability and sampling of nonlinear response.

Models for the Cortisol Profile

For subject i on day j, we measure the salivary cortisol

$$y_{ijk} = f(t_k, \theta_{ij}) + \epsilon_{ijk}$$

- $\epsilon_{ijk} \sim N(0, \sigma^2)$: independent measurement error
- $\theta_{ij} | \theta_i \overset{iid}{\sim} MVN(\theta_i, \Sigma^d)$: parameter specific to subject i on day j
- $\theta_i \overset{iid}{\sim} MVN(\theta, \Sigma^s)$: parameter specific to subject i
- Piecewise Linear Model (Hajat et al., 2010):

$$f(t;\theta) = \theta_0 + \theta_1 t + \theta_2 (t - 0.5h)_+ + \theta_3 (t - 2h)_+$$

- Nonlinear Model (Stroud et al., 2004):

$$f(t;\theta) = \theta_0 + \theta_1 t + \theta_2 t \exp(-\theta_3 t)$$

Cortisol Features

Features	G	References
Baseline	$f(0, \theta)$	Kumari et al. (2009)
Cortisol Awakening Response	$f(0.5, \theta) - f(0, \theta)$	Pruessner et al. (1997)
Evening Decline	$\frac{1}{16-10}(f(16, \theta) - f(10, \theta)dt$	Adam (2006)
Area Under the Curve	$\int_0^{16} f(t, \theta)dt$	Badrick et al. (2007)
Prediction	$f(t, \theta)$	Powell et al. (2002)

Table: Cortisol Features

Optimal Design to Estimate Cortisol Features

Identify the optimal design (n, m, d, T) that minimizes

- the variance of a single cortisol feature
- the weighted sum of variances of several cortisol features

Components of the optimal design (n, m, d, T):

- n: the total number of subjects
- m: the number of days for sampling
- d: the number of samples per day
- $T = (t_1, \ldots, t_d)$: the daily sampling schedule

Existing Approaches

There have been some research on the optimal schedule in the context of PD/PK:

Retout et al. (2002)

- Maximizes the determinant of the information matrix of the parameters
- It is the reciprocal of the size of the confidence region.

Stroud et al. (2001)

- Minimizes the pre-posterior prediction error
- Employs a Bayesian adaptive strategy for the sampling schedule given previous data.

We will take a different perspective:

- Focus on the variance of estimating the cortisol features.
- Analyze the implication of between-subject and between-day variability on multi-level sampling.

Computing the Variance of Cortisol Feature G

Denote the MLE of G by $\hat{G}$ and the information matrix by $I(\theta)$

- $\sqrt{n}(\hat{G} - G) \rightarrow N(0, \nabla G' I^{-1}(\theta) \nabla G)$ by the delta method
- We can show that under some conditions

$$I(\theta) = \nabla f(T,\theta)' \Sigma \nabla f(T,\theta)$$

where

$$\Sigma = \begin{pmatrix} \Sigma^s + \Sigma^d & \Sigma^s & \cdots & \Sigma^s \\ \Sigma^s & \Sigma^s + \Sigma^d & \ddots & \vdots \\ \vdots & \ddots & \ddots & \Sigma^s \\ \Sigma^s & \cdots & \Sigma^s & \Sigma^s + \Sigma^d \end{pmatrix}$$

- Roy et al. (2007) shows that

$$I(\theta)^{-1} = \Sigma^s + \frac{1}{m}(\Sigma^d + \sigma^2 (X'(T)X(T))^{-1})$$

Computing the Variance of Cortisol Feature G

$$Var(\hat{G}) = \frac{1}{n}\nabla G' I(\theta) \nabla G$$

$$= \underbrace{\frac{\nabla G' \Sigma^s \nabla G}{n}}_{\text{subject}} + \underbrace{\frac{\nabla G' \Sigma^d \nabla G}{nm}}_{\text{day}} + \underbrace{\frac{\sigma^2 \nabla G'(X'(T)X(T))^{-1} \nabla G'}{nm}}_{\text{schedule within day}}$$

Clear Interpretation for each component

- $\nabla G' \Sigma^s \nabla G$: the between subject variability for G
- $\nabla G' \Sigma^d \nabla G$: the between day variability for G
- $\sigma^2 \nabla G'(X'(T)X(T))^{-1} \nabla G'$: the estimation variance assuming no between-subject / between-day variability

Properties of Optimal Design

Optimal schedule, T

- minimizes $\nabla G'(X'(T)X(T))^{-1}\nabla G'$
- No knowledge of the variability components is required.

Optimal number of days, m

- When total number of samples is fixed, more subjects (fewer days) is almost always more efficient.

$$Var(\hat{G}(\alpha n, m, d, T)) \leq Var(\hat{G}(n, \alpha m, d, T)) \quad " = " \text{ if } \nabla G' \Sigma^s \nabla G = 0$$

- When the total cost is fixed, the optimal m depends on the cost ratio and within-between variability ratio.

Bayesian Design

- The optimal design depends on unknown parameters.
- Incorporate the uncertainty by considering expected variance (Atkinson et al., 2007)
- Assuming θ and $\Sigma^s, \Sigma^d, \sigma^2$ are independent

$$
\begin{aligned}
E(Var(\hat{G})) &= \int Var(\hat{G}) dp(\theta, \Sigma^s, \Sigma^d, \sigma^2) \\
&= \text{tr}(E(\nabla G' \nabla G)(\frac{E(\Sigma^s)}{n} + \frac{E(\Sigma^d)}{n \cdot m})) \\
&\quad + \frac{E(\sigma^2)}{n \cdot m} E(\nabla G'(X'(T)X(T))^{-1} \nabla G)
\end{aligned}
$$

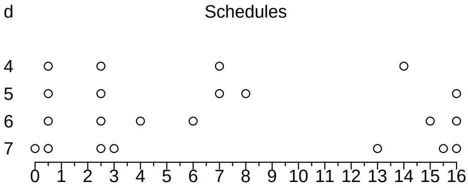
- Simple to compute
- Results on the previous slide still hold
- Only the mean, not the full distribution of $\Sigma^s, \Sigma^d, \sigma^2$ is required.

Application to MESA Stress Study

Context: Determine the optimal design for 2nd-stage data collection.

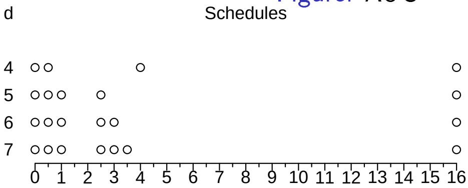
- The estimates and prior distributions are generated from the existing MESA data.
- # of samples per day: $d = 4..7$
- # of days: $m = 1..5$
- Times points of the daily schedule T are chosen from $\{0, 0.5, \ldots, 16\}$

Optimal Design



	Variance				
d	1 Day	2 Day	3 Day	4 Day	5 Day
4	82.7	67.6	62.6	60	58.6
5	76.6	64.6	60.6	58.6	57.4
6	73.4	63	59.5	57.8	56.7
7	71.4	62	58.8	57.2	56.3

Figure: AUC



	Variance				
d	1 Day	2 Day	3 Day	4 Day	5 Day
4	0.562	0.281	0.188	0.141	0.113
5	0.419	0.212	0.140	0.106	0.0854
6	0.369	0.183	0.122	0.0935	0.0738
7	0.333	0.169	0.112	0.0847	0.0676

Figure: CAR

- $\nabla G' \Sigma^s \nabla G = 0.00033 \approx 0$ for CAR
- Expanding the days of sampling is almost as good as expanding subject.

Optimal m for a Fixed Total Cost

- Depends on the cost ratio $\frac{\text{Initial}}{\text{Day}}$ and variability ratio $\frac{\text{Day}}{\text{Subject}}$.

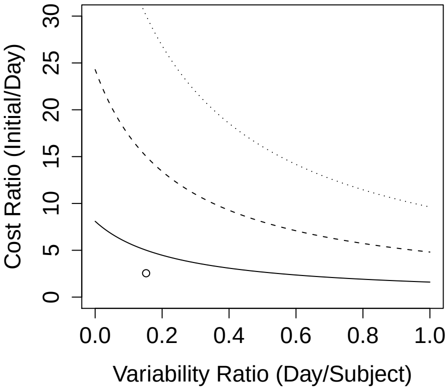


Figure: AUC

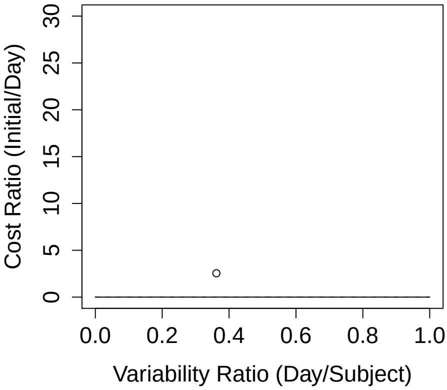
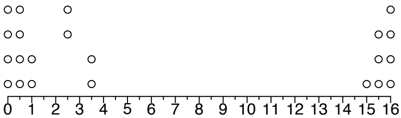


Figure: CAR

Compound Objective

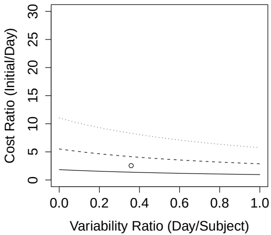
- Minimize the weighted average of variance of Baseline, CAR, Evening Decline and AUC

Schedules



Variance 10^–3

d	1 Day	2 Day	3 Day	4 Day	5 Day
4	3.94	2.62	2.18	1.96	1.83
5	3.45	2.38	2.02	1.84	1.73
6	3.18	2.24	1.93	1.77	1.68
7	3.01	2.15	1.87	1.73	1.64



Summary and Future Work

Summary

- We discussed the properties of the optimal design for longitudinal study.
- The key is the closed form solution of $Var(\hat{G})$.
- We applied these results to MESA Stress Study.
- The results also hold for arbitrary levels of variabilities.

Future Work

- More flexible ways to model variabilities.
- Optimal designs for joint modeling of salivary cortisol and CVD outcome.

Thank you!

Reference

Adam, E. K. (2006). Transactions among adolescent trait and state emotion and diurnal and momentary cortisol activity in naturalistic settings. *Psychoneuroendocrinology*, 31(5):664–679.

Atkinson, A. C., Donev, A. N., and Tobias, R. (2007). *Optimum experimental designs, with SAS*. Oxford University Press.

Badrick, E., Kirschbaum, C., and Kumari, M. (2007). The relationship between smoking status and cortisol secretion. *J Clin Endocrinol Metab*, 92(3):819–824.

Hajat, A., Diez-Roux, A., Franklin, T. G., Seeman, T., Shrager, S., Ranjit, N., Castro, C., Watson, K., Sanchez, B., and Kirschbaum, C. (2010). Socioeconomic and race/ethnic differences in daily salivary cortisol profiles: the multi-ethnic study of atherosclerosis. *Psychoneuroendocrinology*, 35(6):932–943. PMID: 20116177.

Kumari, M., Badrick, E., Chandola, T., Adam, E. K., Stafford, M., Marmot, M. G., Kirschbaum, C., and Kivimaki, M. (2009). Cortisol secretion and fatigue: associations in a community based cohort. *Psychoneuroendocrinology*, 34(10):1476–1485. PMID: 19497676.

Powell, L. H., Lovallo, W. R., Matthews, K. A., Meyer, P., Midgley, A. R., Baum, A., Stone, A. A., Underwood, L., McCann, J. J., Herro, K. J., and Ory, M. G. (2002). Physiologic markers of chronic stress in premenopausal, Middle-Aged women. *Psychosom Med*, 64(3):502–509.

Pruessner, J. C., Gaab, J., Hellhammer, D. H., Lintz, D., Schommer, N., and Kirschbaum, C. (1997). Increasing correlations between personality traits and cortisol stress responses obtained by data aggregation. *Psychoneuroendocrinology*, 22(8):615–625.

Retout, S., Mentré, F., and Bruno, R. (2002). Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Statistics in Medicine*, 21(18):2623–2639.

Roy, A., Bhaumik, D. K., Aryal, S., and Gibbons, R. D. (2007). Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics*, 63(3):699–707. PMID: 17825003.

Stroud, J. R., Müller, P., and Rosner, G. L. (2001). Optimal sampling times in population pharmacokinetic studies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):345–359.

Stroud, L. R., Papandonatos, G. D., Williamson, D. E., and Dahl, R. E. (2004). Applying a nonlinear regression model to characterize cortisol responses to corticotropin-releasing hormone challenge. *Annals of the New York Academy of Sciences*, 1032:264–266. PMID: 15677424.