Applied Statistics at AVC:
How to deal with Repeated Measures Data

Henrik Stryhn

CVER - Centre for Veterinary Epidemiological Research
Atlantic Veterinary College
University of Prince Edward Island, Canada

Graduate Studies and Research Days 2013
Zoetis Lecture

Introduction — Aim

Repeated Measures (Measurements) or Longitudinal Data:

- a series of measurements/records on the same "unit" or "subject" (e.g., individual, well or farm), usually over time,
- allows exploration of effects over time of subjects and "treatments" applied to them,
- commonly encountered data structure in research at AVC.

Analysis of Repeated Measures Data:

- wide selection of methods available (and vast literature),
- discussed in some AVC graduate courses (VHM 802 + 831/832),
- aims of presentation:

  to outline some guiding principles for analysis, and to illustrate how exploration of repeated measures data may lead to interesting insights into the biological systems involved.

Types of Repeated Measures Data

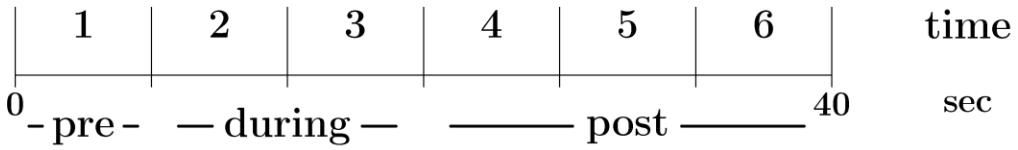
1) Continuous (quantitative, interval-scale) measurements:
   - involves models based on normal distribution,
   - primary focus of this presentation,
2) Discrete (categorical, binary or count) records:
   - similar to (1) but require very different analytical methods when normal distributions cannot be assumed,
3) Survival for subjects observed repeatedly over time:
   $\Rightarrow$ survival analysis (totally different analytical approaches),

What about time series analysis? — main (rough) distinction:

- time series: single or a few long series of measures, analyzed to determine patterns over time,
- repeated measures: many series of measures, analyzed to compare developments over time between subjects.

Example: Clinical Study on Heart Rates of Dogs

- Researcher: Dr. Etienne Côté (Dept. Companion Animals),
- Objective: to investigate the effects of physical manipulations[1] by a veterinarian (Dr. Côté) on the heart rates (among other things) of dogs:
  - application of ocular pressure (OP) and control/sham (OP-),
  - carotid sinus massage (CSM) and control/sham (CSM-).
- Design: 32 dogs randomly distributed to 4 treatment groups, monitored by electrocardiogram for approx. 40 secs, divided into 6 measurement intervals in which average heart rates (HR) were computed:



[1] So-called vagal maneuvers, referring to the vagus nerve.

Why is Repeated Measures Data / Analysis Special? (a bit)

We cannot use "standard" or usual methods:

- because the multiple measures on the same subject are dependent (not independent, as assumed by "standard" methods),
- the number of "units" for treatments (applied to subjects) equals the number of subjects, not the number of measurements.

The study focus or objective usually involves time, e.g. comparison of treatment groups over time,

- effects/comparisons rarely expected to be constant over time,[2]
- can also study dynamics over time for the subjects (contrary to a design with different subjects at each time point),
- the choice of time points in design and analysis is important,
  - regular time points across subjects makes the data easier to analyze and interpret.

[2] In statistical terms, we have interactions with time.

Exploring the Data Graphically

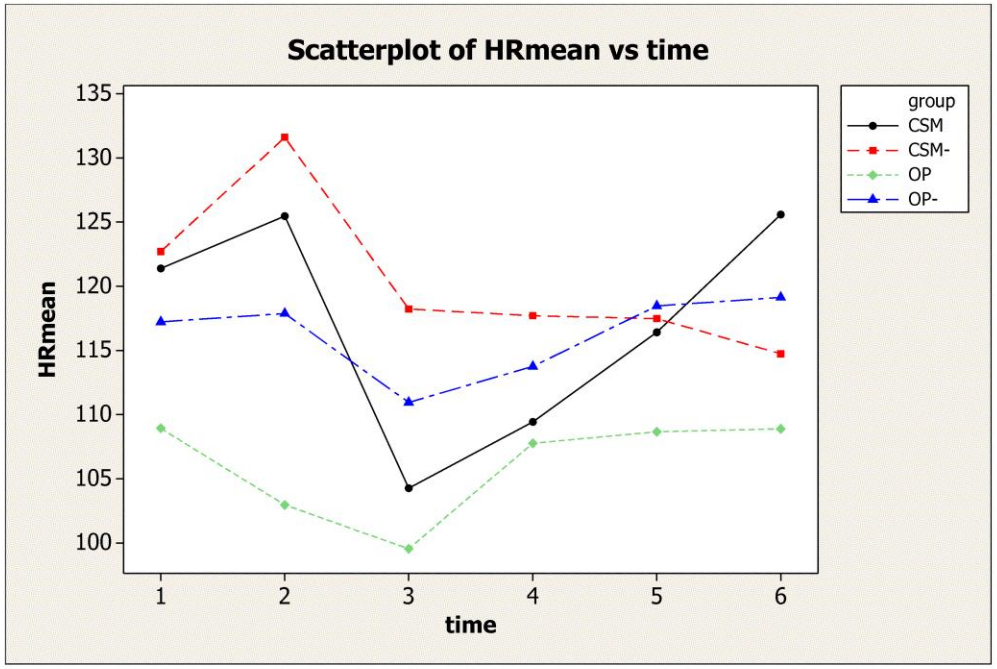
Two main types of plots against time, both recommended:

- Mean plot: average values across "groups" of subjects (e.g. for treatments),
  - shows group trends over time,
  - commonly used for presentation of results.
- Profile plot[3]: series of values over time for subjects,
  - shows variability between subjects and consistency of mean patterns in "groups" of subjects,
  - pick suitable subsample(s) if the dataset is too large to plot all profiles in one graph.

Insights from graphical exploration are crucial to guide the statistical analysis, in particular the choice of analytical approach.

[3] The name "spaghetti plot" is also used (Hedeker and Gibbons, 2006).

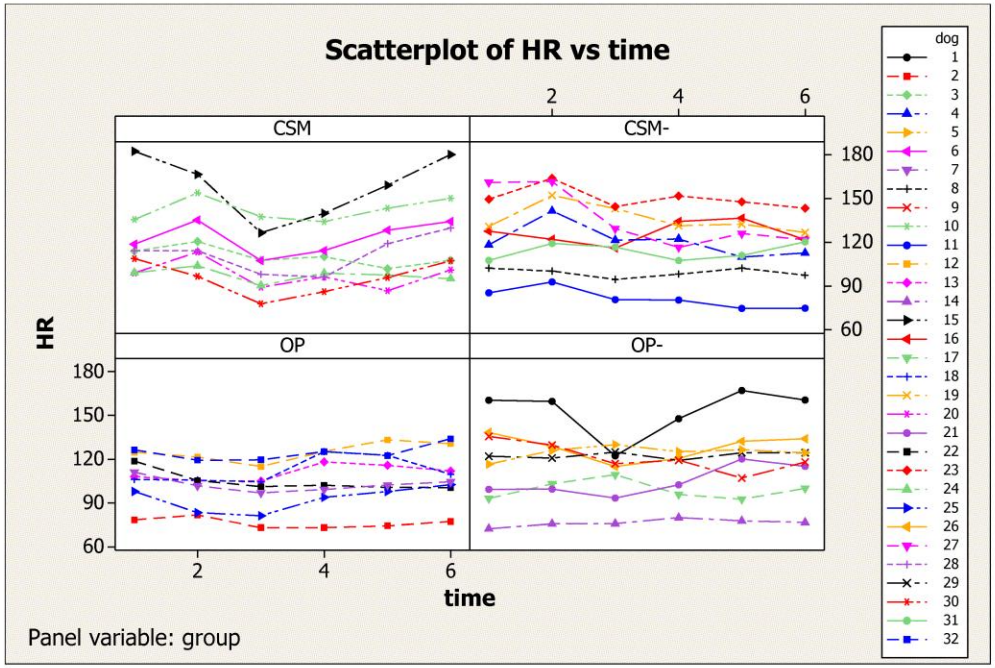
Heart Rate Data: Mean Plot



Interpretations:

- variable curves over time for the 4 groups,
- some pre-treatment group differences (perhaps not significant).

Heart Rate Data: Profile Plot



- strong differences between dogs in their HR levels,
- variable patterns over time across dogs, but no obvious outliers,
- average time trends from mean plot hardly visible here.

A First Simple Approach: Analysis at Separate Time Points

Idea: analyse for some (biologically motivated) or all time points with respect to the design of the subjects (e.g. treatment groups).

Advantages:

- eliminates the repeated measures $\Rightarrow$ simple analysis,
- corresponds to information shown (vertically) in the mean plot.

Drawbacks:

- does not include development over time in statistical analysis,
- involves multiple tests (correlated because on the same subjects) $\Rightarrow$ correction for multiple testing[4] necessary.

Summary: limited analysis with low power, but valid and may be sufficient to demonstrate effects at the selected time point(s).

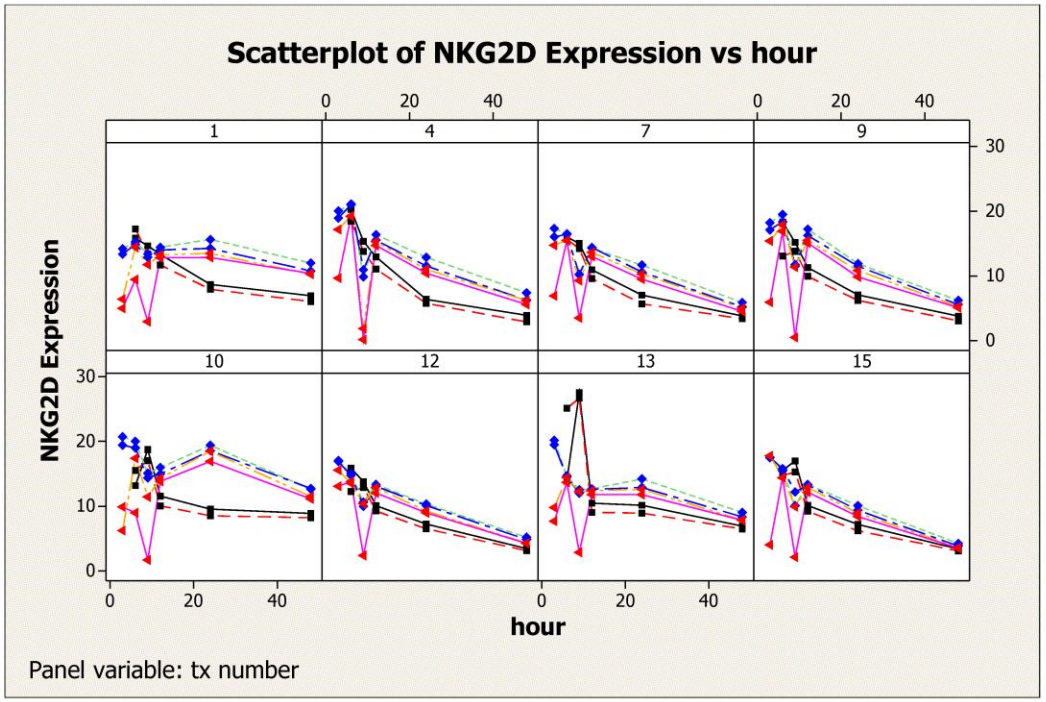
[4] One option is a Bonferroni correction by which all P-values are multiplied by the number of time points analysed.

Example: In-Vitro Assays on Effects of Cytokines on Cell Activity

- Researchers: PhD Candidate Gailene Tobin, Dr. Michael van der Heuvel (both Dept. Biomedical Sciences),
- Objectives: (overall) to determine the effect of certain cytokines (IL–12 and IL–18) on the activation of natural killer (NK) cell immune response, (specifically) to quantify temporal effects on NKG2D receptor expression when adding different concentrations of IL–12 and IL–18 to NK–92 cells,
- Design (subset): in each of 3 trials, NK–92 cells in 16 wells organized in 2 rows were subjected to 8 different concentrations of IL–12 and IL–18 $(0 - 100\, ng/mL)$, in addition to a baseline concentration of IL–2 $(10\, ng/mL)$, and NKG2D expression was quantified by flow cytometry[5] of samples extracted from each well at (0), 3, 6, 9, 12, 24 and 48 hours after treatment.

[5] Response variable: difference between geometric means for stained and isotype control.

Data: NKG2D Expression Assays



Plot symbols/colors ∼ trials 1–3.

Note lower values and slightly different patterns in trial 1.

- apparently noisy initial reaction followed by gradual development over time (mostly a decline),
- 24 hours chosen as a suitable outcome time, confirmed by separate cell viability testing.

Results: NKG2D Expression Assays

Statistical analysis for 24 hours (without trial 1):

- ANOVA: strongly significant ($P < 0.001$) effects of treatments, trials and rows (within plates), but no significant interactions,
- estimated treatment means (SE = **0.14**) with letter codings (a–f) to indicate significant differences (Tukey method), displayed in the table $\sim$ incomplete factorial design:

Mean (NKG2D)	IL–18 dose			
IL–12 dose	0	10	50	100
0	14.1^{b}	18.4^{a}	–	12.9^{c}
10	11.5^{d}	11.0^{de}	–	–
50	–	–	9.7^{f}	–
100	10.5^{e}	–	–	9.2^{f}

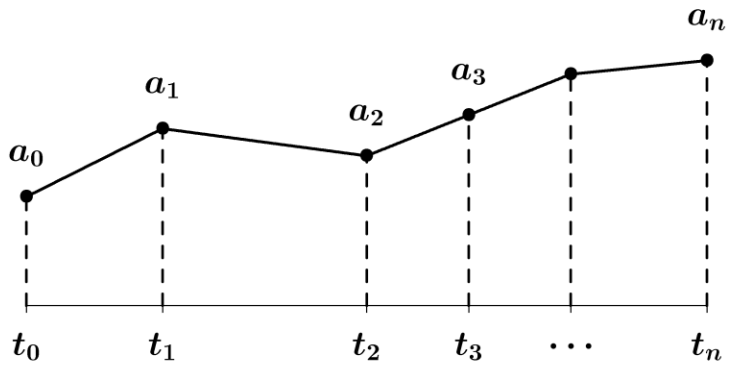
Conclusion: Analysis at 24 hours produced useful results (without utilizing all the data).

A Second Simple Approach: Area under the Curve

Idea (Summary Statistic): compute from each subject's profile a single value that characterizes the profile in a meaningful way and has potential to distinguish between subjects,

- example: gain (last value minus first value),
- example: area under the curve (AUC), interpretable as the accumulation of values through the time period.

Calculation of AUC by the trapezoidal rule/formula:



$$\text{AUC} = (t_1 - t_0)\frac{a_0 + a_1}{2} + (t_2 - t_1)\frac{a_1 + a_2}{2} + \ldots + (t_n - t_{n-1})\frac{a_{n-1} + a_n}{2}$$

If the time points $t_0, \ldots, t_n$ are equidistant:

$$\text{AUC} = (t_1 - t_0)(a_0 + \tfrac{1}{2}(a_1 + \ldots + a_{n-1}) + a_n)$$

Example: Cow Nutrition Pilot Field Trial in Kenya

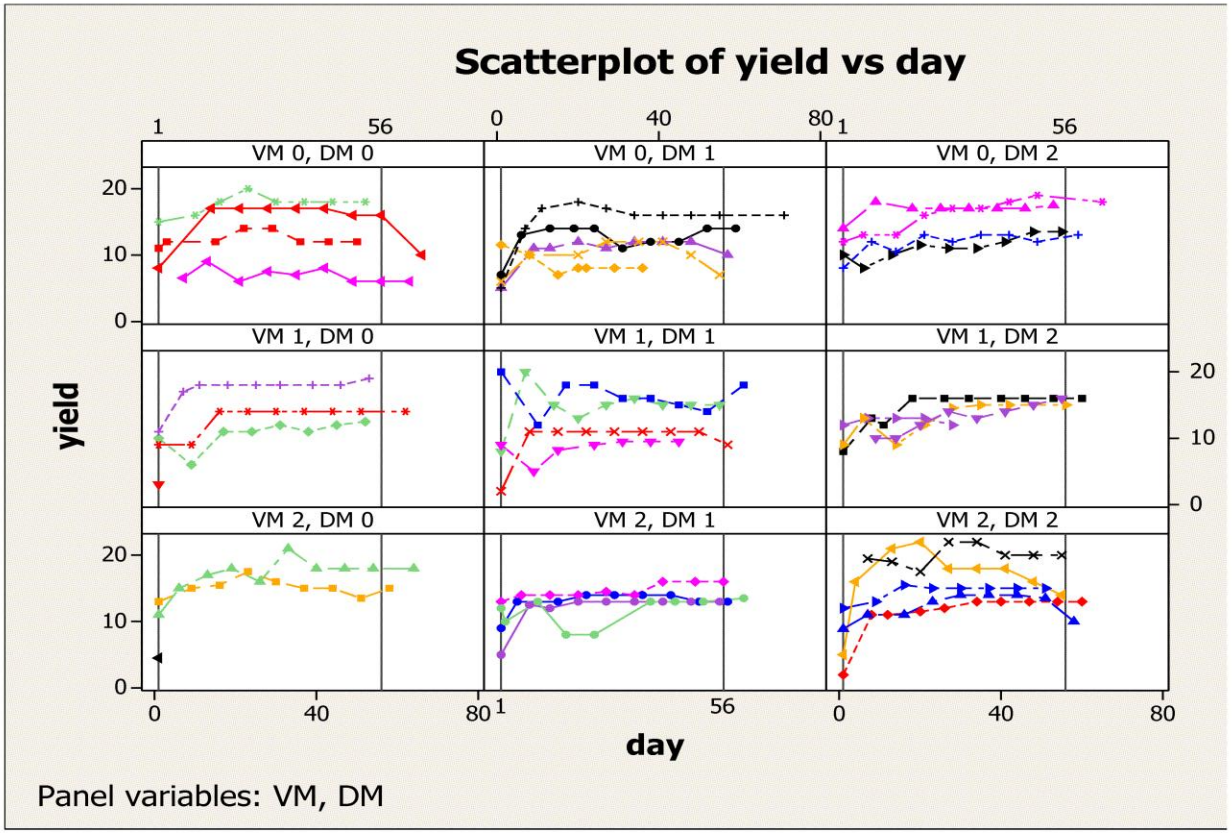
- Researchers: Drs. Shauna Richards & John Vanleeuwen (Dept. Health Management), Sylvia Situma (Univ. of Nairobi, Kenya),
- Objective: to investigate the effects of 9 different cow supplementation diets on milk yield after calving in smallholder dairy farms in Kenya:

Vitamin & Mineral (VM) Supplement	Dairy Meal (DM) Supplement		
	none(0)	half(1)	full(2)
none(0)	4 cows	5 cows	4 cows
half(1)	3 cows[a]	4 cows	4 cows
full(2)	2 cows[a]	4 cows	5 cows

[a] excluding one cow that died shortly after calving

- Design: 35 cows randomly distributed to the 9 diets, which were maintained for 10 weeks, with approx. weekly farm visits and measures of milk yield.

Data: Cow Field Trial in Kenya



- large between-cow variability but no obvious effects of diets.

Results: Cow Field Trial in Kenya

Illustration of AUC method:

- computation not straightforward with irregular time points: AUC computed for days 1–56, with missing endpoint values estimated from the data,
- AUC $\sim$ estimated total milk yield during period (days 1–56).

Results for AUC method:

- descriptive statistics: mean $\mathbf{742}\,\boldsymbol{kg}$, range $\mathbf{391 - 1103}\,\boldsymbol{kg}$,
- ANOVA model (DM,VM,DM*VM) adjusting also for initial body condition (BCS0, values: $\mathbf{1.5, 2, 2.5, 3}$):
  - large unexplained variation: $\boldsymbol{s = 172}\,\boldsymbol{kg}$, $\boldsymbol{R^2 = 19.6\%}$,
  - no significance for DM,VM,DM*VM,BCS0 (all $P \gg 0.10$).

Conclusion: no effect of diet groups on total milk yield 1–56 days.

Overview: Complex Approaches for Full Data Analysis

Issues to deal with in analysis of full repeated measures data set:

- accounting for the within-subject correlation structure[6] (how correlations depend on time: details on next slide),
- choice of appropriate analysis scale to meet model assumptions (e.g. normal distributions for errors),
- assessment and statistical exploration of any time-dependent effects (i.e. time interactions)[6],
- incorporation of unequal variances as needed (commonly across the time points),
- choice of statistical software (major limitations in many software packages, least so in: R, SAS, Stata).

Recommendation: a full analysis of repeated measures data requires some relevant training and/or help (guidance).

[6] The regularity of time points determines the options and difficulty of this step.

Correlation Structure and Correlation Matrices

Correlation structure (CS):

- 2 measures on same subject are correlated (dependent), but this correlation (ρ) generally depends on the corresp. time points:
  close in time $\Rightarrow$ (maybe) high ρ,
  far away in time $\Rightarrow$ (maybe) low ρ,
- correlations among $(Y_1, \ldots, Y_n)$ are displayed in an $n \times n$ matrix $C(Y)$, e.g. for $n = 4$:

$$C(Y) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{21} & 1 & \rho_{23} & \rho_{24} \\ \rho_{31} & \rho_{32} & 1 & \rho_{34} \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 \end{pmatrix}$$

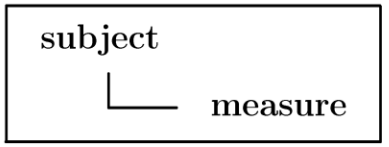
Idea: include CS (for the errors) in both model specification and inference,

- some common, "intuitive" structures for regularly spaced data:

$$\left\{ \begin{array}{c} \text{"equal"}, \\ \text{comp.} \\ \text{symm.} \end{array} \right\} \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho & \rho & 1 & \\ \rho & \rho & \rho & 1 \end{pmatrix}, \left\{ \begin{array}{c} \text{auto-} \\ \text{reg.,} \\ \text{ar(1)} \end{array} \right\} \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \rho^2 & \rho & 1 & \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}, \left\{ \begin{array}{c} \text{Toep-} \\ \text{litz} \end{array} \right\} \begin{pmatrix} 1 & & & \\ \rho_1 & 1 & & \\ \rho_2 & \rho_1 & 1 & \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix}$$

Two Limited Approaches for Full Data Analysis

Random effects model (i.e., subject random effects, "split-plot approach") $\sim$ hierarchical data structure:



- implicitly assumes "equal" (compound symmetry) CS — unnatural and may be give poor data fit for long series,
- possibly ok for short series or with little unexplained variation.

Repeated measures ANOVA (for regular time points):

- adjustment for the CS of $\boldsymbol{F}$-tests in random effects model ANOVA table for effects involving time,[7]
- approximation method with uncertain power and major drawbacks:
  - only subjects with complete data series can be included,
  - does not provide any post-ANOVA inference (appropriate SE, CI or pairwise comparisons).

[7] Precisely, an adjustment for deviations in CS from the minimal assumption required to make these $\boldsymbol{F}$-tests valid (Davis, 2002).

Mixed Models[8] for Repeated Measures Data

- includes correlation structure $C(\varepsilon)$ of specified type(s) for the within-subject errors (ε), in addition to usual fixed (and possibly also random) effects,
- statistical estimation and inference based on the likelihood function[9] (which strongly limits the choice of software[10]),
  - nested models for CS can be tested against each other by likelihood-ratio tests,
  - models for CS can be ordered by model selection criteria such as the AIC,
- recommendation: simple, or parsimonious, modelling of CS is preferable, as fixed effects inference can be affected by CS.

[8] As these models do not always have random effects, the name Covariance Pattern Models is also used (Hedeker and Gibbons, 2006).

[9] The likelihood function may be thought of as goodness-of-fit statistic, and the maximum likelihood estimates are chosen to maximize the (full or restricted) likelihood function.

[10] Facilities exist in R (nlme library), SAS (mixed procedure), Stata (xtmixed command) and SPSS, but differ in their flexibility (e.g., built-in choices of CSs).

CS Modelling in Practice: Heart Rate Data

Model settings (for demonstration purposes):

- analysis on (natural) log scale, effects: Group,Time,Group*Time,
- $-2 \log L$ and AIC (smaller is better) for a range of CS models:

Corr. structure	# params.	$-2 \log L$	AIC
comp. symm.	2	-278.3	-274.3
ar(1)	2	-298.7	-294.7
arma(1,1)	3	-298.9	-292.9
ar(1)+rand.eff.	3	-308.8	-302.8
Toeplitz	6	-318.7	-306.7
ar(1)+het.var.	7	-302.8	-288.8
ar(1)+r.e.+het.	8	-315.8	-299.8
Toeplitz+het.	11	-321.4	-299.4
unstructured	21	-333.8	-291.8

ar(1)+rand.eff.:
$\sigma^2(\varepsilon) = 0.0063$
$\sigma^2(\text{dog}) = 0.0341$
$\rho = .577$ ($\rho_1 = .934, \ldots$)

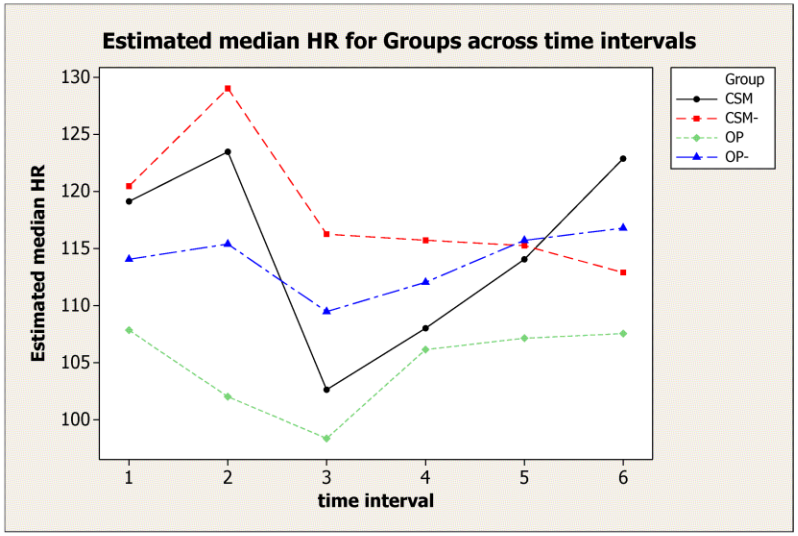
Toeplitz:
$\sigma^2(\varepsilon) = 0.039$
$(\rho_1, \rho_2, \ldots, \rho_5) =$
$(.931, .870, .856, .876, .905)$

- large between-dog variability $\Rightarrow$ high within-dog correlations,
- within-dog correlations decline slowly with time distance,
- variance heterogeneity across time not a problem.

Dealing with Time Effects: Heart Rate Data

Outline of Presentation and Interpretation of time effects:

- Toeplitz CS model: strong sign. ($P < .001$) for Time & Group*Time,
- relevant graphical tool is the interaction plot, e.g. backtransformed ($v \mapsto \exp(v) = e^v$) to yield estimated medians:
- typical questions of interest:
  - mean and SE?: possible, but median and CI easier
  - group diff. at time points? here non-sign. at all times (note low sample size of 8)
  - time diff. within groups? here all sign. except OP– (pairwise comp. possible)

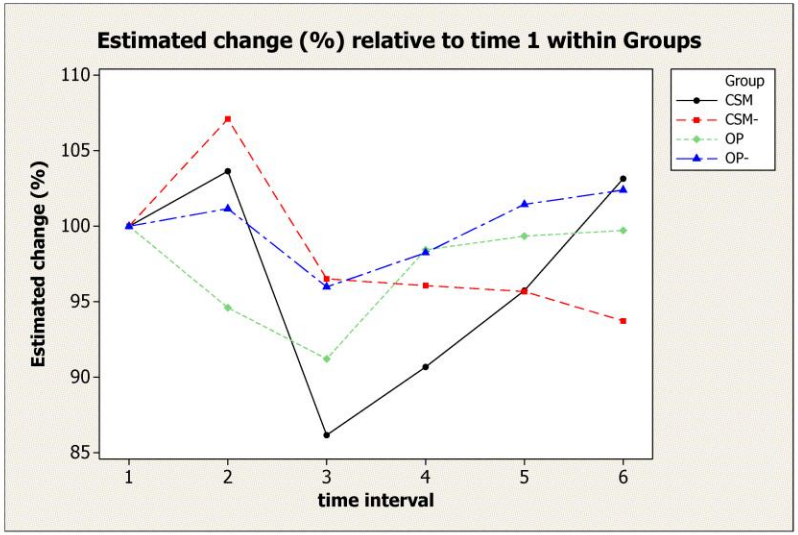


- usually not of interest: overall comparison of groups (disregarding time), or full pairwise comparisons for Group*Time.

Dealing with Time Effects II: Heart Rate Data

Alternative idea for presentation of results of analysis on log-scale:

- additive model on log-scale $\sim$ multiplicative model on original scale $\Rightarrow$ parameter estimates can be presented as relative change,
- for categorical predictors, a baseline needs to be set, here we chose the initial time interval,
- e.g., for CSM, heart rate at time 3 dropped to 86% of the value at time 1.



Other ideas (Heart Rate data):

- restrict comparisons to manip. types (OP vs. OP–, CSM vs. CSM–),
- estimate manipulation type contrasts over time.

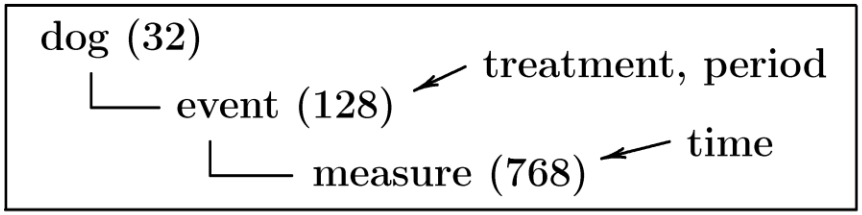
Heart Rate Data: Cross-Over Repeated Measures Trial

Full design/data includes all 4 manipulation groups for each dog,

- cross-over trial where each subject goes through all treatments,[11]
- order of treatments events (called periods 1–4) for each dog determined by Latin square design[12], e.g. for four of the dogs (with treatments denoted by A–D):

Treatment	Period			
Dog no.	1	2	3	4
1	B	D	A	C
2	A	B	C	D
3	C	A	D	B
4	D	C	B	A

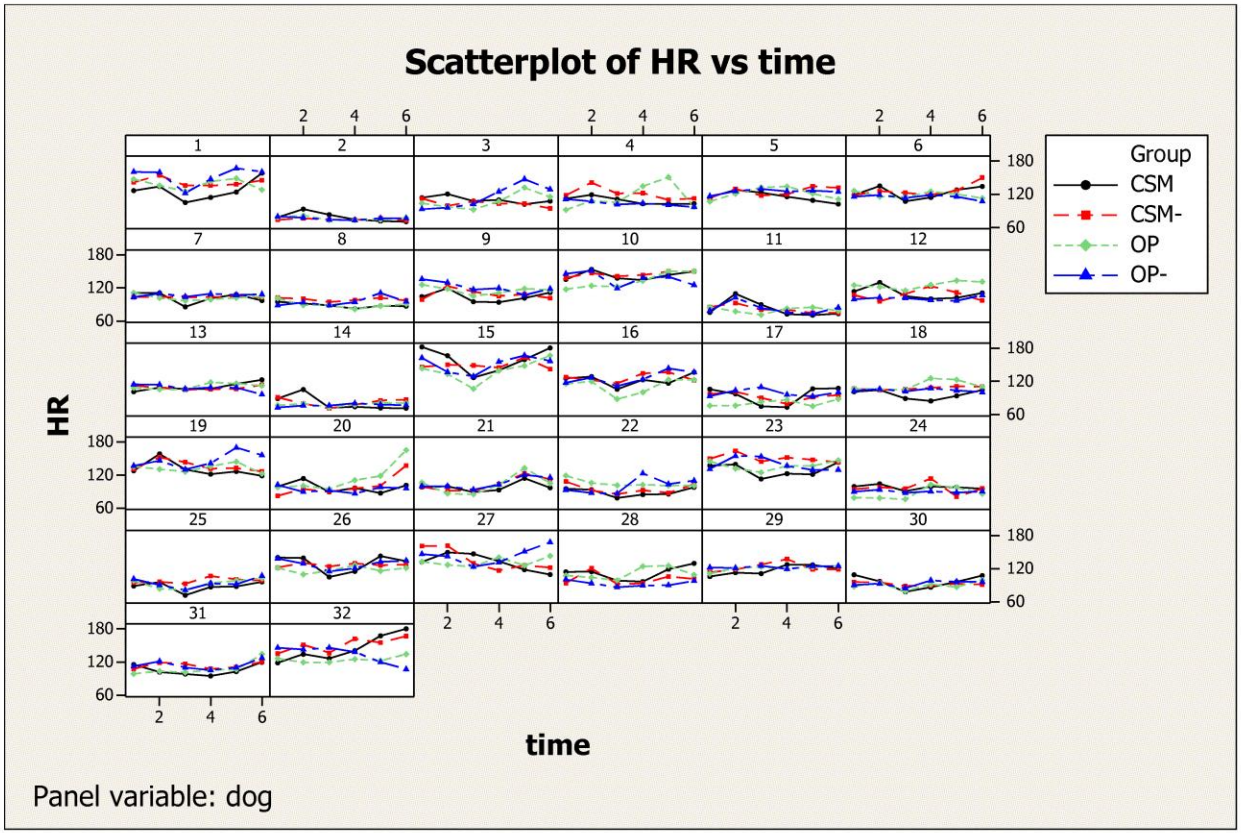
Full data structure: more complex, and may be viewed as hierarchical:



[11] Cross-over trials can substantially improve power and scope because treatments are now compared within (instead of between) subjects, provided that no carry-over effects occur.

[12] A Latin square design is characterized by each treatment occuring once in every row and column, leading to balancedness in two blocking factors, here the dogs and periods.

Data: Full Cross-Over Trial



- larger differences in HR levels between dogs than across groups within the same dog ⇒ cross-over design a clear improvement.

Mixed Model for Full Heart Rate Data

Natural starting point for analysis:

- keep previous modelling of Group*Time and within-event CS,
- add random effects for dogs, and period effect + interactions.

Some modifications of the approach required/proposed:

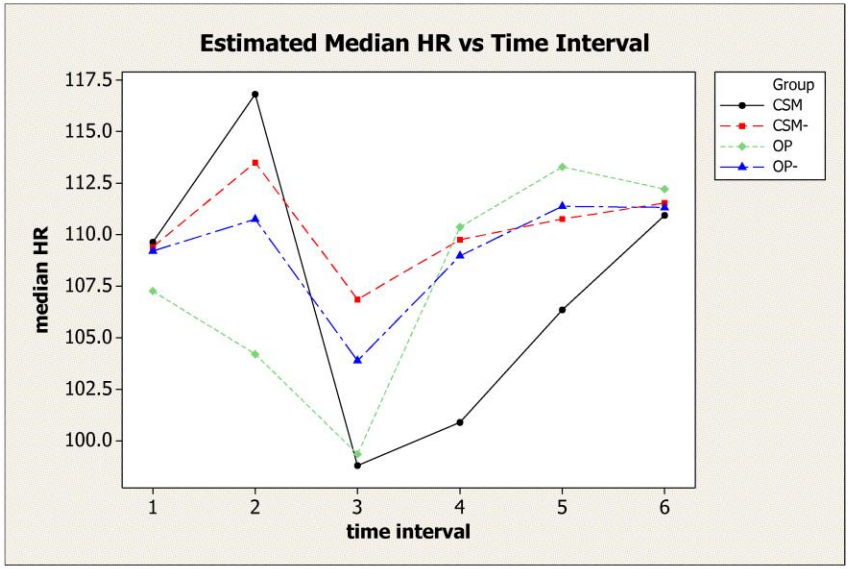
* compliance with model assumptions: log-transform with offset[13],
* correlation structure for total of 24 measures per dog:
  may not be sensible with same correlations between any 2 measures from different events, so try:
  - multivariate correlation structures[14],
  - random slopes (group/period effects may vary between dogs),
  - non-constant variances across times or periods.

[13] Box-Cox analysis for transformation power and offset led to suggested log-transform with an offset of (-25), i.e.: $\text{HR} \mapsto \ln(\text{HR}-25)$.

[14] A limited selection of multivariate correlation structures are available in SAS, proc mixed.

Results: Group*Time for Full Heart Rate Data

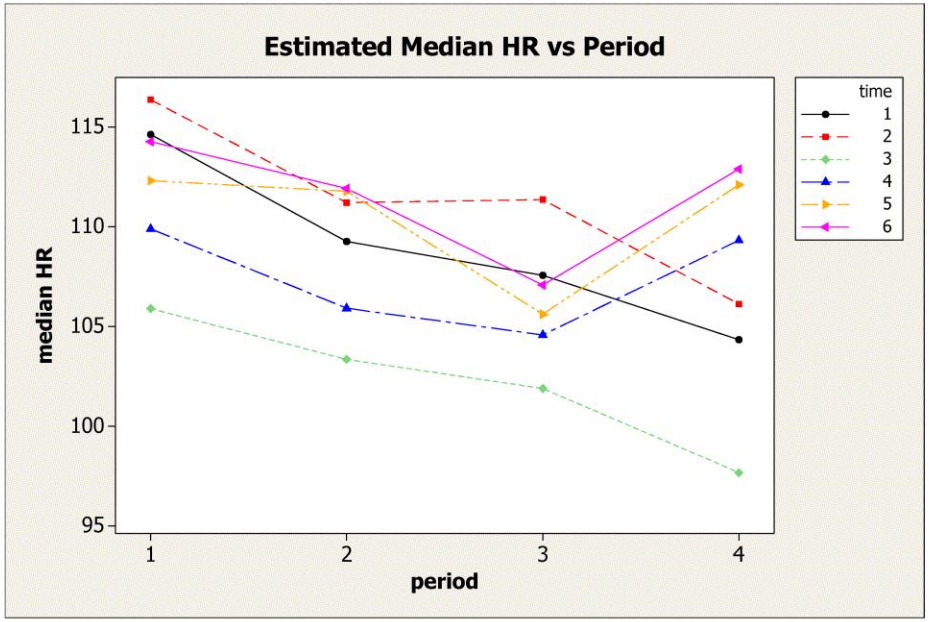
- estimated median HR (SE $\sim$ **3.4**$-$**4.2**):
- statist. significance:
  - Group*Time ($P < .001$)
  - across times for all groups ($P < .004$)
  - CSM vs. time 1: times 2($\uparrow$), 3($\downarrow$), 4($\downarrow$), adj. for CSM$-$: times 3($\downarrow$), 4($\downarrow$)
  - OP vs. time 1: times 3($\downarrow$), 5($\uparrow$), 6($\uparrow$), but no significance after adjustment for OP$-$,



- interpretation (main finding): drop in HR for CSM during and immediately post manipulation (also when compared to CSM$-$), but return to baseline level afterwards.

Results: Period*Time for Full Heart Rate Data

- estimated median HR (SE $\sim$ **3.3**$-$**4.2**):



- stat. significance: moderate for Period*Time ($P = .009$), strong at times 1–3 ($P < .003$), weak for times 4–6 ($P = .02-.07$),
- interpretation: declining HR through study ($\sim$ acclimation effect), except for rise in last period, for recovery intervals only.

Results: Correlation Structure for Full Heart Rate Data

Exploration of complex variance and correlation structures:

- no indication of random slopes for period or dog-specific group effects,
- "un@ar(1)" multivariate CS for periods and time[15] with 12 parameters gave best model fit (by AIC), but showed only weak heterogeneity in variances and between-event correlations,
- best parsimonious CS was arma(1,1) within events combined with dog random effects (no additional variation at events):

$$\sigma^2(\text{dog}) = 0.0539,\ \sigma^2(\varepsilon) = 0.0135,\ \rho = .224,\ \gamma = .443,$$

  - excl. dog effects: moderate within-event ρ's (.443 ↓ .001),
  - incl. dog effects: 80% of unexplained variance for dogs; high within-event ρ's (.889 ↓ .800); cross-event ρ (ICC) of .800,
- interpretation: very large between-dog variation and low additional within-event correlation (mostly one time step).

[15] The formula implies an ar(1) within-event CS and unstructured correlations and variance heterogeneity among the four periods; the resulting correlations for pairs of measures is obtained by multiplying the respective terms.

How to Deal with Repeated Measures Data: Take-Home Message

1) In our planning, explicitly discuss how the objective and expectations for the study involve time effects/dynamics, and let our study design (e.g., choice of time points) reflect that.

2) Explore our data graphically, to get a picture of effects of and over time,

3) Use information from 1) and 2) to select one (or several) method(s) for analysis that match our ambition level and analytical capability,

   - a simple method may be fine if it answers our questions and we are aware of its limitations,
   - if a complex analysis is "necessary", we may need help to execute and interpret it.

4) Include in our analysis always: (i) evaluation of its assumptions, and (ii) detailed presentation and discussion of results (in particular for effects involving time).

5) Write up the study/analysis using relevant reporting guidelines (e.g., the SAMPL guideline for Statistical analyses and methods; equator-network.org).

A Few Final Remarks

Conclusion:
Repeated measures data offer lots of opportunities for interesting and challenging exploration and analysis . . . (and perhaps fun).

Acknowledgements — my thanks go to:

- the students and faculty who supplied data and insight for this lecture,
- my colleagues who generously (an understatement) nominated me for the award,
- all collaborators — students, staff, and faculty — who invite me into their projects and provide inspiration for (statistical) thought and research,
- the audience — Thank You for Your Attention!

Some Suggestions for Further Reading

The literature on repeated measures is voluminous and overwhelming; as a start, you may want to consult the course notes for VHM 802 and Chapter 23 of the VER textbook (Dohoo et al., 2009).

Recommended textbooks on Repeated Measures Analysis:

Davis, C. R., 2002, *Statistical Methods for the Analysis of Repeated Measurements*. Springer.

Dohoo, I. R., Martin, S. W. and Stryhn, H., 2009. *Veterinary Epidemiologic Research*, 2nd ed. VER-Inc., Charlottetown, Canada; Chapter 23.

Hedeker, D. and Gibbons, R. D., 2006. *Longitudinal Data Analysis*. Wiley.

Pinheiro, J. C. and Bates, D. M., 2000. *Mixed-Effects Models in S and S-PLUS*. Springer; Chapters 1–5. *(A bit technical)*

Rabe-Hesketh, S. and Skrondal, A., 2012. *Multilevel and Longitudinal Modeling using Stata*, 3rd ed. Stata Press; Chapters 5–6.

Singer, J. D. and Willett, J. B. 2003, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press; Chapters 1–7.