

MCG Dept. of Biostatistics
Journal Club
September 28, 2009

Stephen W. Looney

ARTICLE

Matthews JNS, Altman DG, Campbell MJ,
Royston P. Analysis of serial
measurements in medical research.
British Medical Journal 1990; 300: 230-5.

STATEMENT OF THE PROBLEM

How does one decide what method to
use when analyzing serial (repeated)
measurements?

- Single repeated (within) factor
- Single grouping (between) factor

Examples from Personal Consulting Experience

- Compare two treatments (denavir & oil
of lavender) for herpes simplex lesions
- Examine association of blood lead
levels with systolic & diastolic bp
- Perform repeated measures analysis
with two grouping factors & one
repeated factor.

Serial Measurements

➤ A common study design in medical
research is to give patients an
intervention and then observe what
happens to them over time.

- Example 1 (Matthews et al.)
 - ❖ Compare aspirin absorption in patients who
are either healthy or ill.
 - ❖ Each patient was given the same dose of
aspirin per kg body weight & his or her blood
aspirin concentration was measured at time
zero & after 5, 10, 15, 20, 30, 40, 60, 75, 90, and
120 min.

“Usual” Method of Analysis (Fig. 2)

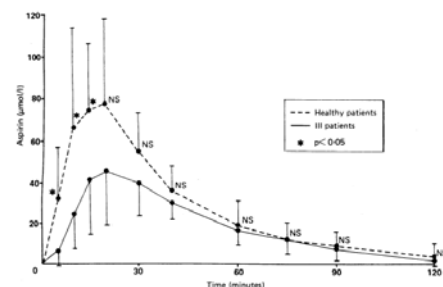
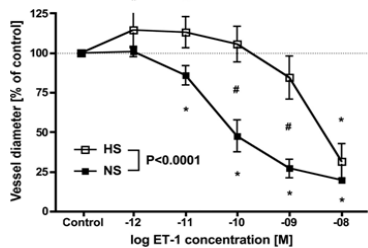


FIG 2—Mean and standard deviation of aspirin concentrations in nine healthy and nine ill patients over time.
("Usual" method of display)

- Compare 2 groups separately at each time pt.

Example 2 (“Rightward Shift”)

Fig. 2. Response of afferent arterioles from rats on NS (n = 6) and HS (n = 6) diets, expressed as percentage of control diameter, to increasing concentrations of ET-1



Schnieder, M. D. et al. Am J Physiol Renal Physiol 292: F1209-F1214, 2007. doi: 10.1152/ajprenal.00290.2006

AJP - Renal Physiology

Types of Time Dependency

➤ Peaked

❑ In many studies the outcome variable starts from a baseline (sometimes zero), rises to a peak, and then returns to baseline.

❖ This is displayed as a peaked curve (Fig 1).

Purpose of Matthews et al.

➤ “To propose a general simple method for a clinically useful and statistically valid analysis.”

❑ “Method of Summary Measures”

❑ They consider only studies in which each patient receives a single treatment or intervention

❖ Excludes escalating dose studies & crossover trials

❑ Their focus is on continuous data, but their method could also be applied to ordinal data.

Types of Time Dependency

➤ Growth

❑ Sometimes the outcome variable steadily increases or decreases with time and does not return to its initial value over the period of study.

❖ This is displayed as a growth curve (Fig. 1)

Figure 1

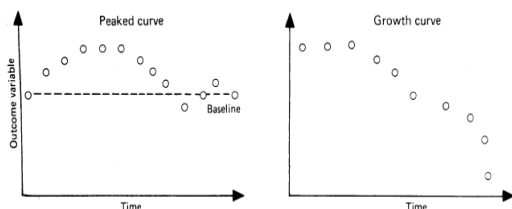


FIG 1—Examples of peaked curve and growth curve

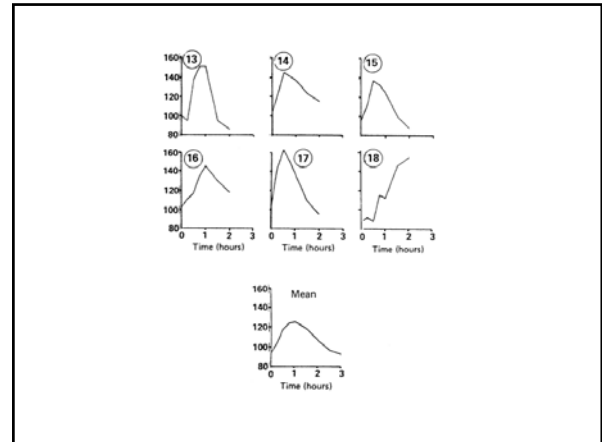
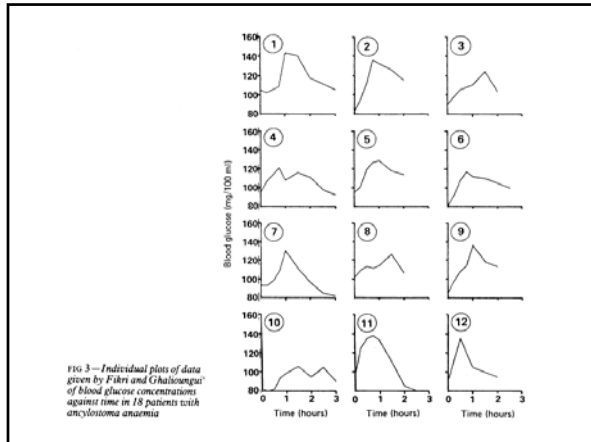
Problems with “Usual Method”

(1) The curve joining the group means may not be a good descriptor of a typical curve for an individual.

❑ Important variation in the shapes and locations of curves for different subjects may be hidden.

❖ Figure 3 shows individual glucose tolerance curves for 18 subjects with ancylostoma anemia.

❖ It is evident that the mean curve hides a wide variety of curves, including multiple peaks (cases 4 and 10) and a steady rise (case 18).



Problems

(2) No account is taken of the fact that measurements at different time points are from the same subject.

It is inherent in the design that the *main interest lies in the way individual subjects respond over time*, yet this is ignored when each time point is analyzed separately.

Problems

(3) Successive observations on a given subject are likely to be correlated

The value at one time point is likely to be associated with those at successive time points, so the corresponding significance tests will not be independent.

If a test at one point in time gives a significant result, then it is likely that tests performed at points close in time will also give significant results.

Problems, cont.

➤ Points (1) and (2) together indicate that the usual analysis may give a misleading impression of the way that individual subjects typically respond over time and give no information about variation among subjects in their response over time.

➤ The error bars relate only to between subject variation at each time point.

Problems, cont.

➤ From point (3) it is evident that there are convincing statistical arguments against multiple tests.

Separate significance tests at different time points are often interpreted as if they gave independent information about the relative location of the groups.

Problems, cont.

- ❑ Dividing the results into "significant" and "not significant" introduces an artificial dichotomy into serial data.
- ❑ Most biological variables change over time in a smooth and continuous manner, and so the idea that at one point in time the difference between two variables is not significant whereas at the next point in time it is significant is artificial.

"Method of Summary Measures"

- First described by Wishart (*Biometrika* 1938)
 - ❑ Treats the individual as the basic unit of analysis and uses responses for each subject to construct a single number which summarizes some aspect of that subject's response curve.
 - ❑ *This approach avoids all the difficulties outlined above.*

Some Summary Measures (Appendix I)

Type of data	Question to be answered	Summary measure
Peaked	Is the overall value of the outcome variable the same in different groups?	Overall mean (equal time intervals) Area under curve (unequal time intervals)
Peaked	Is the maximum (minimum) response different between groups?	Maximum (minimum) value
Peaked	Is the time to maximum (minimum) response different between groups?	Time to maximum (minimum) response

Summary Measures, cont.

Type of data	Question to be answered	Summary measure
Growth	Is the rate of change of the outcome variable different between groups?	Regression coefficient
Growth	Is the eventual value of the outcome variable the same between groups?	Final value of outcome measure or difference between last and first values, or percentage change between first and last
Growth	Is the response in one group delayed relative to the other?	Time to reach a particular value (for example, a fixed percentage of baseline)

Personal Experiences

- Compare two treatments (denavir & oil of lavender) for herpes simplex lesions
 - ❑ Summary Measures Used
 - ❖ Elapsed time from beginning of treatment until pain score of 0 was obtained
 - ❖ Reduction in size of lesion (Day 4 – Day 1)
 - ❖ Rate of change in pain and size of lesion
 - Slope of regression of pain or size on elapsed time (22 time pts between baseline & Day 4)

Personal Experiences, cont.

- Examine association of blood lead levels with systolic & diastolic bp
 - ❑ Each variable was measured at baseline, 6, 18, 30, 42, 54 mos
 - ❑ Summary Measures Used
 - ❖ Area under curve for blood lead and bp
 - ❖ Rate of change in blood lead and bp
 - Slope of regression line for each one regressed on time

Personal Experiences, cont.

- Repeated measures analysis with two grouping factors & one repeated factor
 - ❑ Too many interactions to deal with
 - ❑ Summary Measures Used in ET-1 Proposal
 - ❖ Area under curve for vessel diameter
 - ❖ Maximum value of smooth muscle signaling response
 - Perform standard two-way ANOVA using AUC or maximum as dependent variable

Peaked Response Curves

- If the outcome measure is the concentration of a substance that has been administered (e.g., aspirin, glucose) the response will often be peaked.
 - ❑ Total uptake of the substance may be of interest and can be measured by the area under the response curve in an individual subject.
 - ❖ Mean of the observations can also be used as a measure of the overall value.

Peaked Response Curves, cont.

- Another feature of peaked data that is frequently of interest is the maximum value.
 - ❑ May be interpreted as a measure related to the maximum effect of the agent given.
- The time taken to reach the maximum (or minimum) may also be a clinically important variable.

Growth Curves

- Rate at which the variable is changing over the duration of the experiment is usually an important feature.
 - ❑ A good measure of this rate will often be the slope of the line fitted to the data using least squares.
- The “final outcome,” possibly expressed as a difference from baseline, may usefully represent some “achievable” value.
 - ❑ This may be more appropriate for data that tend to level off over the last few values.

Choice of Summary Measure

- More than one summary measure can be used so that different aspects of the response may be investigated.
 - ❑ No general rule for how many measures should be used.
 - ❑ Each measure should aim at summarizing a different aspect of the response.
 - ❑ “There are seldom more than two or three interesting aspects of a response curve....”

Choice of Summary Measure, cont.

- The summary measures should have some clear clinical or biological relevance.
- They should be chosen before the data are collected.
 - ❑ This avoids the temptation to choose a particular summary measure because it shows a maximal difference between groups.
 - ❑ *“It would be dogmatic, however, to insist that a summary measure could not be chosen after the data have been examined, especially when the shape of the time-response curve is uncertain.”*

Choice of Summary Measure, cont.

➤ Thinking in terms of relevant summary measures may enable the clinical investigator to formulate specific research questions that can be addressed once the data are obtained.

❑ By thinking in these terms in advance it may be possible to improve the design of the study.

❖ For example, if the important summary measure is time to maximum response, then more frequent measurements would be needed around the time that the maximum response is expected.

Analysis Issues

➤ Once the appropriate summary measure has been calculated for each subject, its values can be treated as raw data for an appropriate statistical analysis.

❑ Most common: group comparisons of means or medians of the summary measures.

➤ “There are few computer packages that allow computation of summary measures in a routine manner.”

❑ Mean(X_1, X_2, \dots, X_P)

❑ Max(X_1, X_2, \dots, X_P)

❑ Use SAS macro with PROC REG to calculate regression slopes for all subjects & then output into single data set.

Graphical Display of Serial Measurements

➤ “The most informative approach” is to produce separate graphs of the responses against time for each subject.

❑ Same axis scaling should be used on all graphs.

❑ Example: Figure 3, p. 231

Graphical Display, cont.

➤ The plots for each subject can be arranged into a panel or grid with separate panels for each group.

➤ It may help to order the plots in some way, such as by increasing mean or maximum value.

Graphical Display, cont.

➤ “Spaghetti Plot” can be used to display response profiles for all subjects on the same set of axes.

❑ Example – Figure 4, p. 233

❖ Aspirin Absorption Data

Figure 4, Healthy Patients

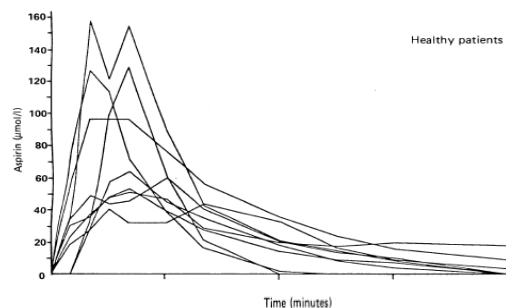


Figure 4, Ill Patients

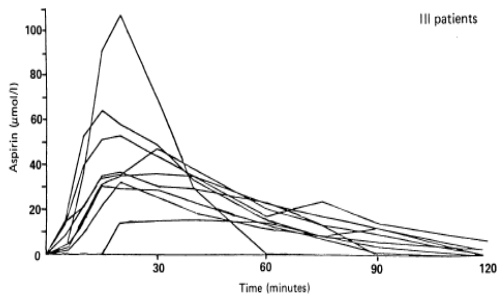
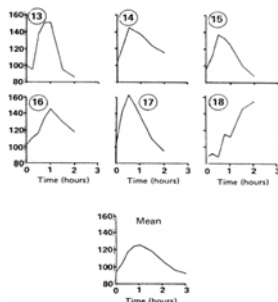


FIG 4—Individual plots of aspirin concentrations against time in healthy patients and ill patients

Graphical Display, cont.

- “Mean curve”
 - ☐ Response averaged over all subjects and then plotted vs. time
 - ☐ Most likely to be useful when peak response occurs at about the same time for all subjects
 - ❖ e.g., when subjects all respond quickly to a stimulus.

Figure 3



Graphical Display, cont.

- Summary measures have considerable advantages for plotting because simple graphical methods, such as histograms, scatter plots, etc., can be used.
- Insight can be gained from scatterplots of two summary measures against each other.
 - ☐ e.g., plot the maximum (or minimum) value for each subject against the time that the maximum (or minimum) occurred.
 - ❖ Example: Figure 5, p. 233 (aspirin absorption data)

Figure 5

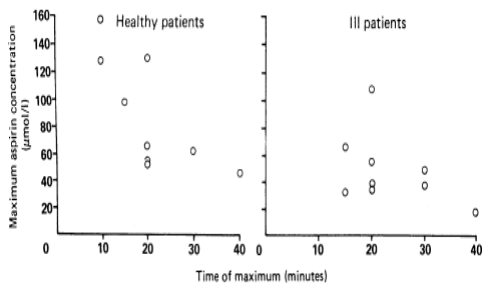


FIG 5—Scatter plot of maximum aspirin concentrations by time of maximum in healthy patients and ill patients

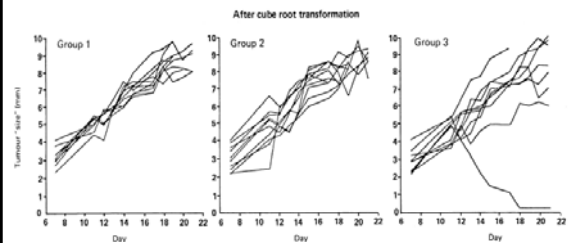
Conclusions from Fig. 5

- Clearly shows that peak values tend to be lower and occur later in the ill patients.
- There also seems to be a negative relation between the size of the peak and its time of occurrence
 - ☐ Higher peaks occur earlier, particularly in healthy patients
 - ☐ This is not easily seen in Fig 4.

Growth Curve Example

- Changes in tumor volume in three groups of 10 rats at 11 time points, after different injections.
 - ❑ Tissue culture medium (Group 1)
 - ❑ Tissue culture medium and normal spleen cells (Group 2)
 - ❑ Normal spleen cells, immune RNA, and tumor antigen (Group 3)

Figure 6, p. 234 (bottom panel)



- Growth rates appear more linear after transformation.
- One obvious outlier in Group 3 (tumor regression).
- Summarize data for each animal by regressing tumor size on day.

Growth Curve Example, cont.

- One-way ANOVA (excluding the outlier) shows no significant difference in the rate of growth of tumors in the three groups.
 - ❑ Mean tumor growth rates = 0.438, 0.438, and 0.435 mm/day for groups 1, 2, and 3 respectively, p -value ≈ 1.000 .
- Note that the estimated slopes for each rat were used as the “data” for purposes of analysis.

Discussion

- “We have shown that there are serious problems associated with the common use of comparisons at each time point when analyzing serial measurements on patients.
 - ❑ The method is inappropriate because it does not provide clear answers to clinically relevant questions.
 - ❑ There are important statistical deficiencies.”

Discussion, cont.

- “We have described the method of summary measures, which avoids all of the problems identified with the more usual analysis.”
 - ❑ Its main disadvantage is that it may be difficult to specify in advance an appropriate summary measure.
 - ❑ The use of this type of analysis should encourage researchers to think about the features of the data that will be of most interest to them when designing the study.

Discussion, cont.

- If there are missing values, as is common with this type of study, the method of summary measures will usually still allow the summary to be calculated, but other methods may be difficult to apply.
- If the times of observations differ among subjects, then all the methods, other than summary measures, will fail.

Disadvantages of Method of Summary Measures

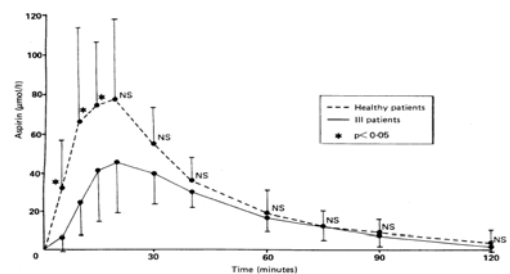
- The method tacitly assumes that the same general pattern (peaked, growth, etc.) is present in all groups being compared.
 - ❑ The same summary measure must be used in all treatment groups.
 - ❑ This assumption may be violated, especially if one of the groups being compared is a negative control group.

Disadvantages, cont.

- Some research questions that may be of interest to the investigator cannot be addressed using this method:
 - ❑ How does the outcome variable change over time?
 - ❖ We may wish to compare the mean response at Time 1 with the mean at Time 2, etc.
 - ❖ Or, we may wish to model the *trend* in mean response (linear, quadratic, etc.)

Disadvantages, cont.

- ❑ Does the overall pattern in the outcome variable differ over time between groups?
 - ❖ We wish to determine if there is significant *interaction* between the grouping variable and the repeated factor.
 - Sometimes, we wish to determine the time point at which the two groups first begin to differ.



- Is the overall pattern the same between groups?
- Do the mean responses differ between time pts?

Alternative Methods of Analysis

➤ Traditional Repeated Measures Analysis (Looney & Stanley, *American Statistician*, 1989)

- ❑ Test for interaction between "group" & "time"
 - ❖ If not significant, test "group" & "time" main effects
 - Test of group main effect is equivalent to using mean as the summary measure in Matthews et al.
 - ❖ If significant, test "group" & "time" simple effects at each level of other factor using Bonferroni or similar adjustment

Alternative Methods of Analysis

- Traditional Repeated Measures Analysis
 - ❑ PROC GLM
 - ❑ Requires complete data at all time points
 - ❑ Compromise between multivariate & adjusted univariate approaches (i.e., do both using $\alpha/2$ as significance level)
 - Mixed-effects Models (PROC MIXED)
 - ❑ Easily handles missing data
 - ❑ Can consider many possible covariance structures (autoregressive, compound symmetry, etc.)
 - Random Effects Regression (RRM) Models
 - ❑ Allows modeling of individual subjects' responses over time

Conclusion

➤ The method of summary measures can be a useful adjunct to standard methods for analyzing repeated measures data.

❑ However, its limitations with regard to addressing certain relevant research questions should be acknowledged.

❑ Practicing biostatisticians should make sure both methods are in their statistical toolkit.

Thank you!