Methods for analysing cardiovascular studies with repeated measures

T.J. Cleophas, A.H. Zwinderman, B.M. van Ouwerkerk

Background. Repeated measurements in a single subject are generally more similar than un-repeated measurements in different subjects. Un-repeated analyses of repeated data cause underestimation of the treatment effects.

Objective. To review methods adequate for the analysis of cardiovascular studies with repeated measures.

Results. (1) For between-subjects comparisons, summary measures and random-effects mixed-linear models are possible. Examples of summary measures include the area under the curve of drug time-concentration and time-efficacy curves, maximal values, mean values, and changes from baseline. A problem is that precision is lost because averages, rather than individual data, are applied. Random-effects mixed-linear models, available in SPSS statistical software and other software programmes, provide better precision for that purpose. (2) For within-subjects comparisons, repeated-measures ANOVAs are available in SPSS and other software programmes. Subgroup factors such as gender differences and age class can be included.

Discussion. For non-Gaussian data, Wilcoxon's and Friedman's tests are available, for binary data McNemar's tests can be used in case of two repeated observations. No standard methods are available for repeated binary measures with more

than two observations. The purpose of this review was not to present a complete report but, rather, to underline that ample efforts should be made to account for the special nature of repeated measures. (Neth Heart J 2009;17:429-33.)

Keywords: repeated-measures analysis-of-variance (ANOVA), random-effects mixed-linear models

Repeated measurements in a single subject are generally more similar to one another than un-repeated measurements in different subjects. Figure 1 shows a crossover study of the effect of two vasodilators on arterial flow in ten patients. All patients are tested twice. There is a strong positive correlation between the two observations in one patient: the patients who respond well the first time, tend to do so the second time. If the difference in efficacy between the two vasodilators is analysed using a two-samples t-test, then no significant difference is observed ($0.05 < p < 0.10$). If, however, a paired t-test is applied, then a highly significant difference at $p < 0.001$ is obtained. Obviously, if one does not take into account the repeated nature of the data, the treatment efficacy will be underestimated.

Statistical analyses, thus, have to take the similarity of repeated measures in a single subject into account. For that purpose special tests have been designed: paired t-tests, repeated-measures analysis of variance (ANOVA), paired tests for binary data, and paired non-Gaussian data. A general advantage of repeated measures is that less subjects are required for answering scientific questions.

An unpaired analysis of the paired data of a cardiovascular crossover study will no longer be published nowadays, but cardiovascular study designs are increasingly complex and it is common to include repeated observations from a single patient in addition to single observations. The current paper reviews methods for the analysis of such data. Particular attention will be given to (1) the calculation of summary measures, (2) random-effects mixed-linear models, and (3) repeated-measures ANOVA.

T.J. Cleophas

European Interuniversity College of Pharmaceutical Medicine, Lyon, France and Department of Statistics, Circulation, Boston, USA

A.H. Zwinderman

European Interuniversity College of Pharmaceutical Medicine, Lyon, France and Department of Biostatistics, Academic Medical Center, Amsterdam, the Netherlands

B.M. van Ouwerkerk

Department Medicine, Albert Schweitzer Hospital, Dordrecht, the Netherlands

Correspondence to: T.J. Cleophas
Department of Medicine, Albert Schweitzer Hospital,
PO Box 306, 3300 AH, Dordrecht, the Netherlands

Between-subjects comparisons

Calculation of summary measures for overall analysis of repeated measures

It is appropriate, whenever possible, to use a summary estimate of repeated data. For example, the area under the curve of drug concentration-time curves is used in clinical pharmacology as an estimate of bioavailability of a drug.¹ Also, maximal values, mean values, and changes from baseline are applied for the same purpose. The disadvantage of this approach is that it does not use the data fully because summary measures are used instead of the individual data and, therefore, precision may be lost but, otherwise, the approach is unbiased and can be used perfectly well. As an alternative and more precise method a random-effects mixed-linear model, as reviewed below, can be used.

Random-effects mixed-linear models for overall analysis of repeated measures

As an alternative to summary measures, a random-effects mixed-linear model can be used. In table 1 an example is given of a parallel-group study of the effect of two statins on high-density lipoprotein (HDL) cholesterol. HDL cholesterol is measured every week for five weeks. The averages of the five repeated measures in one patient are calculated and an unpaired t-test is used to compare these averages in the two treatment groups. The overall average in group 0 was 1.925 (standard error of the mean (SEM) 0.0025) and in group 1 it was 2.227 (standard error (SE) 0.227). With 18 degrees of freedom and a t-value of 1.99 the difference did not obtain statistical significance, $0.05 < p < 0.10$. As expected, there seems to be a strong positive correlation between the five repeated measurements in one patient. In order to take account of this strong positive correlation, a random-effects mixed-linear model can be used, which is available in SPSS through the commands 'analyse → mixed models'.² For this particular analysis all measurements have to be ordered in a single column, not in five columns side by side. In a second column the time of the separate measurements is noted. We enter the outcome as dependent variable and patients, weeks and treatment modality as factors. Then, we command fixed effects, build nested terms, and model: week, treatment and week*treatment, and ok. We find that after adjustment for the repeated nature of the data, treatment modality is a highly significant predictor of HDL cholesterol levels. Treatment 1 performs significantly better than treatment 0.

Within-subjects comparisons

Repeated-measures ANOVA without between-subjects covariates

Summary measures are, of course, not adequate to assess the differences between the separate measures. The study in table 2 shows that a repeated-measures

Figure 1. Cardiovascular crossover study of the effect of two different vasodilators on arterial flow in ten patients with Raynaud's disease. All patients were tested twice. There is a strong positive correlation between the two observations in one patient: the patients who respond well the first time, do so the second time.

ANOVA can be performed in a sample size as small as four subjects. The study compares the effects of three different treatments to reduce vascular resistance and contains only 12 data. According to the data analysis there is a significant difference between the three treatments with $F=18.2$ and $p=0.024$. In order to assess the appropriateness of the linear assumption of the model a quadratic relationship between indicator and outcome variables is used, but the F value was smaller and, thus, the p value was smaller (11.01 and $p=0.045$) and so the linear assumption seems appropriate. Between-subjects differences are assumed not to influence the treatment comparisons and are, therefore, not taken into account in the analysis.

Repeated-measures ANOVA with between-subjects covariates

In the study in table 3 an example is given of a study where both repeated and non-repeated factors are combined. Three treatment modalities for the treatment of exercise tachycardias are assessed in both male and female subjects of different age classes: 20-30, 30-40, and 40-60 years of age. The variable 1 gives the age class (1, 2, and 3, respectively). The variable 5 gives the genders (1 and 2, respectively).

The variables 2, 3, and 4 give the exercise heart rate during treatment with high-dose, low-dose and very-low-dose β -blocker, respectively, in beats/min. The study tries to answer seven research questions:

1. Does treatment modality influence exercise heart rate?
2. Do subjects from different age classes have different heart rates?
3. Do males have different heart rates than females?
4. Does the pattern of differences between pulse rates for the age class groups change between the genders?
5. Does the pattern of differences between pulse rates for the age class groups change between the treatment modalities?

Table 1. Random-effects mixed-linear model, data file of a parallel group study of 20 patients.

Patient	Week	Outcome	Treatment	Patient	Week	Outcome	Treatment
1,00	1,00	1,66	0,00	11,00	1,00	1,57	1,00
1,00	2,00	1,62	0,00	11,00	2,00	1,82	1,00
1,00	3,00	1,57	0,00	11,00	3,00	1,83	1,00
1,00	4,00	1,52	0,00	11,00	4,00	1,83	1,00
1,00	5,00	1,50	0,00	11,00	5,00	1,82	1,00
2,00	1,00	1,69	0,00	12,00	1,00	1,60	1,00
2,00	2,00	1,71	0,00	12,00	2,00	1,85	1,00
2,00	3,00	1,60	0,00	12,00	3,00	1,89	1,00
2,00	4,00	1,55	0,00	12,00	4,00	1,89	1,00
2,00	5,00	1,56	0,00	12,00	5,00	1,85	1,00
3,00	1,00	1,92	0,00	13,00	1,00	1,83	1,00
3,00	2,00	1,94	0,00	13,00	2,00	2,08	1,00
3,00	3,00	1,83	0,00	13,00	3,00	2,12	1,00
3,00	4,00	1,78	0,00	13,00	4,00	2,12	1,00
3,00	5,00	1,79	0,00	13,00	5,00	2,08	1,00
4,00	1,00	1,95	0,00	14,00	1,00	1,86	1,00
4,00	2,00	1,97	0,00	14,00	2,00	2,11	1,00
4,00	3,00	1,86	0,00	14,00	3,00	2,16	1,00
4,00	4,00	1,81	0,00	14,00	4,00	2,15	1,00
4,00	5,00	1,82	0,00	14,00	5,00	2,11	1,00
5,00	1,00	1,98	0,00	15,00	1,00	2,80	1,00
5,00	2,00	2,00	0,00	15,00	2,00	2,14	1,00
5,00	3,00	1,89	0,00	15,00	3,00	2,19	1,00
5,00	4,00	1,84	0,00	15,00	4,00	2,18	1,00
5,00	5,00	1,85	0,00	15,00	5,00	2,14	1,00
6,00	1,00	2,01	0,00	16,00	1,00	1,92	1,00
6,00	2,00	2,03	0,00	16,00	2,00	2,17	1,00
6,00	3,00	1,92	0,00	16,00	3,00	2,22	1,00
6,00	4,00	1,87	0,00	16,00	4,00	2,21	1,00
6,00	5,00	1,88	0,00	16,00	5,00	2,17	1,00
7,00	1,00	2,04	0,00	17,00	1,00	1,95	1,00
7,00	2,00	2,06	0,00	17,00	2,00	2,20	1,00
7,00	3,00	1,95	0,00	17,00	3,00	2,25	1,00
7,00	4,00	1,90	0,00	17,00	4,00	2,24	1,00
7,00	5,00	1,91	0,00	17,00	5,00	2,20	1,00
8,00	1,00	2,07	0,00	18,00	1,00	1,98	1,00
8,00	2,00	2,09	0,00	18,00	2,00	2,23	1,00
8,00	3,00	1,98	0,00	18,00	3,00	2,28	1,00
8,00	4,00	1,93	0,00	18,00	4,00	2,27	1,00
8,00	5,00	1,94	0,00	18,00	5,00	2,24	1,00
9,00	1,00	2,30	0,00	19,00	1,00	2,21	1,00
9,00	2,00	2,32	0,00	19,00	2,00	2,46	1,00
9,00	3,00	2,21	0,00	19,00	3,00	2,57	1,00
9,00	4,00	2,16	0,00	19,00	4,00	2,51	1,00
9,00	5,00	2,17	0,00	19,00	5,00	2,48	1,00
10,00	1,00	2,36	0,00	20,00	1,00	2,34	1,00
10,00	2,00	2,35	0,00	20,00	2,00	2,51	1,00
10,00	3,00	2,26	0,00	20,00	3,00	2,55	1,00
10,00	4,00	2,23	0,00	20,00	4,00	2,55	1,00
10,00	5,00	2,20	0,00	20,00	5,00	2,52	1,00

Source	Numerator df	Denominator df	F	Significance
Intercept	1	90	6988,626	0,000
Week	4	90	0,377	0,825
Treatment	1	90,000	20,030	0,000
Week * treatment	4	90	1,603	0,181

The outcome is HDL cholesterol levels after treatment for one to five weeks. The SPSS file uses commas instead of dots.

Table 2. Repeated measures ANOVA, effects of three treatments on vascular resistance (blood pressure / cardiac output).

Subject	Treatment 1	Treatment 2	Treatment 3	SD ²	
1	22.2	5.4	10.6	147.95	
2	17.0	6.3	6.2	77.05	
3	14.1	8.5	9.3	18.35	
4	17.0	10.7	12.3	21.4	
Treatment mean	17.58	7.73	9.60		
Grand mean	= 1.63				
SS _{within subject}	= 147.95 + 77.05 +				
SS _{treatment}	= (17.58 – 11.63) ² + (7.73 – 11.63) ² +				
SS _{residual}	= SS _{within subject} – SS _{treatment}				
SPSS statistical software: command: analyse; general linear model; repeated measurements.					
	SS	Dfs	Mean square	F	P value
Within subject	127.2	1	127.2	127.2 / 7.0 = 18.2	0.024
Residual	20.964	3	7.0		

SS=sum of squares, dfs=degrees of freedom equals.

- Does the pattern of differences between pulse rates for genders change between the treatment modalities?
- Does the pattern of differences between pulse rates for treatment modalities change in a subgroup with a particular gender and age class?

The above research questions given in research terms are the following:

- Is there a within-subjects main effect?
- Is there a between-subjects main gender effect?
- Is there a between-subjects main age class effect?
- Is there a between-subjects interaction
- Is there a within-subjects by between-subjects interaction effect?

SPSS statistical software requires the following commands: analyse; general linear model; repeated measurements. According to the data analysis there was a significant difference between the three treatments with $F=2814.9$ and $p=0.000$. In order to assess the appropriateness of the linear assumption of the model a quadratic relationship between indicator and outcome variables was used, but the F value and thus p value were smaller, and, so the linear assumption seems appropriate. It can be hard to interpret the results of interactions. For example, the seventh research question above is confirmed: there is a significant interaction between treatment modality and gender and age class. Averages of the subgroups can be examined in order to understand what is going on (figure 2). Figure 2 shows that gender differences do not remain the same but seem to increase with subsequent treatment modalities. There is, obviously, interaction between treatment modality and gender. However, this is only true in the younger, but not in the older age classes.

Generally, a significant interaction is a disaster for a comparative study of different treatment modalities,

because an overall comparison of the treatment modalities becomes meaningless. In the example given, the magnitude of the interaction is limited, and an overall treatment difference can still be observed (figure 2). The overall result can at least be reported in a qualitative manner.

Discussion

Repeated measures in a single subject are generally more similar to one another than data obtained from entirely different subjects and statistical analyses have to take this into account. For between-subjects comparisons, summary measures or random-effects mixed-linear models are possible, while for within-subjects comparisons repeated-measures ANOVA are available.

Examples of summary measures of repeated measures include the area under the curve of drug time-concentration and time-efficacy curves, maximal

Figure 2. Gender differences do not remain the same, but seem to increase with subsequent treatment modalities. There is interaction between treatment modality and gender. However, this is only true in the younger, not in the older age classes.

Table 3. Repeated measures ANOVA with between-subjects covariates, data file of 54 subjects, the variables VAR 2, 3 and 4 give the individual outcomes after treatment with different compounds inducing exercise tachycardias.

Subject	Age class	VAR 2	VAR 3	VAR 4	Gender	Subject	Age class	VAR 2	VAR 3	VAR 4	Gender
1.	1,00	112,00	166,00	215,00	1,00	28.	2,00	93,00	151,00	217,00	2,00
2.	1,00	111,00	166,00	225,00	1,00	29.	2,00	77,00	122,00	178,00	2,00
3.	1,00	89,00	132,00	189,00	1,00	30.	2,00	78,00	119,00	173,00	2,00
4.	1,00	95,00	134,00	186,00	2,00	31.	3,00	81,00	134,00	205,00	1,00
5.	1,00	66,00	109,00	150,00	2,00	32.	3,00	88,00	133,00	180,00	1,00
6.	1,00	69,00	119,00	177,00	2,00	33.	3,00	88,00	157,00	224,00	1,00
7.	2,00	125,00	177,00	241,00	1,00	34.	3,00	58,00	99,00	131,00	2,00
8.	2,00	85,00	117,00	186,00	1,00	35.	3,00	85,00	132,00	186,00	2,00
9.	2,00	97,00	137,00	185,00	1,00	36.	3,00	78,00	110,00	164,00	2,00
10.	2,00	93,00	151,00	217,00	2,00	37.	1,00	112,00	166,00	215,00	1,00
11.	2,00	77,00	122,00	178,00	2,00	38.	1,00	111,00	166,00	225,00	1,00
12.	2,00	78,00	119,00	173,00	2,00	39.	1,00	89,00	132,00	189,00	1,00
13.	3,00	81,00	134,00	205,00	1,00	40.	1,00	95,00	134,00	186,00	2,00
14.	3,00	88,00	133,00	180,00	1,00	41.	1,00	66,00	109,00	150,00	2,00
15.	3,00	88,00	157,00	224,00	1,00	42.	1,00	69,00	119,00	177,00	2,00
16.	3,00	58,00	99,00	131,00	2,00	43.	2,00	125,00	177,00	241,00	1,00
17.	3,00	85,00	132,00	186,00	2,00	44.	2,00	85,00	117,00	186,00	1,00
18.	3,00	78,00	110,00	164,00	2,00	45.	2,00	97,00	137,00	185,00	1,00
19.	1,00	112,00	166,00	215,00	1,00	46.	2,00	93,00	151,00	217,00	2,00
20.	1,00	111,00	166,00	225,00	1,00	47.	2,00	77,00	122,00	178,00	2,00
21.	1,00	89,00	132,00	189,00	1,00	48.	2,00	78,00	119,00	173,00	2,00
22.	1,00	95,00	134,00	186,00	2,00	49.	3,00	81,00	134,00	205,00	1,00
23.	1,00	66,00	109,00	150,00	2,00	50.	3,00	88,00	133,00	180,00	1,00
24.	1,00	69,00	119,00	177,00	2,00	51.	3,00	88,00	157,00	224,00	1,00
25.	2,00	125,00	177,00	241,00	1,00	52.	3,00	58,00	99,00	131,00	2,00
26.	2,00	85,00	117,00	186,00	1,00	53.	3,00	85,00	132,00	186,00	2,00
27.	2,00	97,00	137,00	185,00	1,00	54.	3,00	78,00	110,00	164,00	2,00

	SS	Dfs	Mean square	F	p value
1. Within-subjects	281916.3	1.35	208537.9	2814.9	0.000
2. Age class	4681.0	2	2340.5	3.045	0.057
3. Gender	26373.0	1	26373.0	34.3	0.000
4. Age class x gender	2155.4	2	1077.7	1.402	0.0256
5. Within-subjects x age class	241.7	2.70	89.4	1.206	0.313
6. Within-subjects x gender	1034.8	1.35	765.4	10.332	0.001
7. Within-subjects x age class x gender	1481.9	2.70	548.1	7.398	0.000

SS=sum of squares, dfs=degrees of freedom. The SPSS file uses commas instead of dots. VAR=variable.

values, mean values, and changes from baseline. A problem with summary measures is that precision is lost, because averages rather than individual data are applied. Random-effects mixed-linear models, available in SPSS statistical software, provide better precision for that purpose.

For within-subjects comparisons, repeated-measures ANOVAs are adequate. They are also available in SPSS as well as other statistical software programmes. Factors such as gender differences and age class can be included in a repeated-measures ANOVA subgroup.

For ANOVA and linear models, the equality of variances and linearity in the data have to be checked. Most software programmes do so routinely, and present alternative approaches if these requirements cannot be satisfied.

We should add that this review is not complete, and there are many other methods. For non-Gaussian data Wilcoxon's and Friedman's tests are available. For binary data McNemar's tests are possible in case of two repeated observations in one subject. Repeated binary measures with more than two observations are rare, and there are no standard methods for evaluation. The main purpose of this review was not to present a complete report but rather to underline that ample efforts should be made to account for the special nature of repeated measures. ■

References

- 1 Benet LZ, Kroetz DL, Sheiner LB. Pharmacokinetic studies. In: Goodman and Gilman's Pharmacologic Basis Therapeutics, 9th Edition, McGraw-Hill, New York, 1997 pp 5-27.
- 2 SPSS Statistical Software 13.0, Chicago, IL, USA, 2000.