Cardiovascular drug trials: how to examine interaction, and why so

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

Background: In practice the benefit of cardiovascular medicines is less consistent than it is in clinical trials. This is due to multiple uncontrolled factors that co-determine the efficacy of the new treatment. In statistical terms, they interact with the new treatment. Interaction effects are rarely assessed in cardiovascular trials.

Objective: To review (1) important factors that may interact with the treatment efficacy, (2) how to examine such factors, and (3) why so.

Results: Important factors include (a) possible risk factors such as specific patient characteristics, and concomitant medications, and (b) study-specific aspects such as heterogeneities of investigators, health centres, and individual patients including patient compliance. Such factors can be assessed by comparing subgroups. A common but incorrect approach is the comparison of the significances of difference between treatment modalities in either subgroup. Instead, a direct comparison of effect sizes relative to the standard errors is adequate. As an alternative, regression modelling is adequate and convenient. Results of interaction assessments

are post-hoc and, therefore, of an exploratory and unconfirmed nature. So, why should they be performed? In cardiovascular research the effects of patient characteristics and drug-drug interactions on drug efficacies are numerous. It is valuable to account at least post-hoc for such mechanisms. Second, current cardiovascular trials involve heterogeneous health centres, investigators, and patient groups. Accounting for these heterogeneities can be helpful to better predict individual responses in future patients.

Conclusion: Cardiovascular trials enrolling patient groups at risk for heterogeneity should include at least a post-hoc assessment for interaction. Correct and incorrect methods for that purpose are described. Interaction assessments are helpful to better predict the efficacy/safety of new cardiovascular medicines in the future treatment of subgroups of patients. (*Neth Heart J* 2007;15:61-6.)

Keywords: research, clinical trials, drug interactions, regression analysis

T.J. Cleophas

European Interuniversity College of Pharmaceutical Medicine, for the EC Socrates Project, Claude Bernard University, Lyon, France and Department of Internal Medicine, Albert Schweitzer Hospital, Dordrecht, the Netherlands

A.H. Zwinderman

European Interuniversity College of Pharmaceutical Medicine, for the EC Socrates Project, Claude Bernard University, Lyon, France and Department of Biostatistics, Academic Medical Centre, Amsterdam, Netherlands

B. van Ouwerkerk

Department of Internal Medicine, Albert Schweitzer Hospital, Dordrecht, the Netherlands

M. Sobh

European Interuniversity College of Pharmaceutical Medicine, for the EC Socrates Project, Claude Bernard University, Lyon, France

Correspondence to: T.J. Cleophas
Department of Medicine, Albert Schweitzer Hospital,
PO Box 444, 3300 AK Dordrecht, the Netherlands
E-mail ajm.cleophas@wxs.nl

In cardiovascular research and drug development, multiple factors such as age, gender, comorbidity, concomitant medication, genetic and environmental factors co-determine the efficacy of the new treatment. In statistical terms, they interact with the new treatment. It is impossible to control for all of these factors. Instead, randomised controlled trials are used to ensure that there are no imbalances between the treatment groups, and an overall assessment is made. The limitation of this approach becomes obvious once the new medicine is applied in practice, where benefits of the new medicines are far less consistent than they are in the trials.¹ Despite this limitation, interaction assessment is rarely included in cardiovascular trials. The current paper was written to review (1) important factors that may interact with the treatment efficacy, (2) how to examine such factors, and (3) why so.

What exactly is interaction?

To illustrate what we mean by interaction, we present a hypothesised example with a patient characteristic as

Table 1. Hypothetical example of a parallel group study (n=400).

	New medicine (n=200)	Control medicine (n=200)	P value
Successfully treated patients	130/200 (65%)	110/200 (55%)	<0.01

Table 2. The same study group as in table 1, including factor gender.

	New medicine (n=200)	Control medicine (n=200)	Add up
Successfully treated females	55/100	65/100	120/200
Successfully treated males	75/100 + 130/200	45/100 + 110/200	120/200

Figure 1. The effect of gender on a treatment comparison of two parallel groups treated with a new and a control medicine. Upper graph: the males respond better to both treatments than the females, and there is no gender $\times$ treatment interaction in the data. Lower graph: the data from the example given in the text: gender $\times$ treatment interaction is possible because the males score better with the new medicine, while the females score better with the control treatment.

interacting factor. The implication of cardiovascular trials of new medicines is, generally, to use the estimated effects in forecasting the results of applying a new medicine to an individual. For that purpose a representative sample of subjects is treated with the new medicine or a control medicine. For example, in a parallel group study 400 patients are treated, as shown in table 1.

Based on this assessment the best bet about the difference between the two treatment modalities is given by the overall difference between the two treatment groups. We can expect the new medicine to perform 10% better than the control medicine. If, however, we include the factor gender into our data, the results look slightly different (table 2).

The results in table 2 show that in the accumulated data, although there is no difference in the numbers of females and males between the two treatments, the new medicine scores better in males, while the control medicine scores better in females. The correct interpretation of these data is that if you do not wish to account for gender, then the new medicine performs better, while if you include only females the control medicine may perform better. The treatment modalities may interact with gender. Interaction effects usually involve situations like this. It is helpful to display interaction effects important to the interpretation of the data in a graph with treatment modality on the x-axis and subgroup results on the y-axis. If the lines drawn for each subgroup are parallel (figure 1, upper graph), there is no interaction in the data. If, however, they have a different slope, and, particularly, if they cross (figure 1, lower graph), then the presence of interaction between treatments and subgroups is suggested, in our example *treatment $\times$ gender interaction*. The new medicine is better in females than it is in males.

The medical concept of interaction is synonymous to the terms heterogeneity and synergism. Interaction must be distinguished from confounding. In a trial with interaction effects the parallel groups have similar

Table 3. Difference between females and males in odds ratios (OR) of treatment success.

	Standard error	p value
OR treatment success females		
0.658	1.336	>0.10
OR treatment success males		
3.667	1.357	<0.001
Difference in OR*		
5.573	2.563	0.05-0.10

*The difference in odds ratios is calculated by subtracting their logarithmic transformations and turning this subtraction sum into its antilogarithmic term.

characteristics. However, there are subsets of patients that have an unusually high or low response. With confounding, things are different. For whatever reason the randomisation has failed and the parallel groups have asymmetric characteristics. For example, in a placebo-controlled trial of two parallel groups asymmetry of age may be a confounder. The control group is significantly older than the treatment group, and this can easily explain the treatment difference. Particularly in survival studies, differences in baseline age may be an important confounder as recently demonstrated by De Craen and Westendorp.²

Incorrect and correct method for testing interaction

In the above example the presence of interaction is suggested. We can statistically test whether the difference between new medicine and control is significantly different using e.g. a χ^2 or an odds ratio test. The tests produce p values of >0.10 for the females and <0.001 for the males. It is tempting to state that the difference in p values establishes a difference between the females and the males. P values are composite estimators of not only effect size, but also spread in the data. Differences in p values can arise because of differences in effect sizes or standard errors.

The correct approach is to compare directly the effect sizes relative to the standard errors, e.g. by using the odds ratios of treatment success in either subgroup (table 3). This procedure produces a p value between 0.05 and 0.10. Obviously, there is a tendency to interaction. However, definitive evidence is lacking.

Using a regression model for testing interaction

Most studies have insufficient size to assess efficacy meaningfully in subgroups of patients. Instead a regression model can be used to evaluate whether the treatment effect interacts with specific covariates. An example is provided by the REGRESS trial, a randomised parallel group trial of 884 patients treated with pravastatin or placebo for two years.³ The primary effect variable was the decrease in the average diameter of the coronary arteries after two years of treatment. The average decrease was 0.057 mm (standard error (SE) 0.013) in the placebo group, and 0.117 mm (SE 0.015) in the pravastatin group (t-test: significance of difference at $p < 0.001$) (figure 2, upper graph). Thus

Figure 2. The effect of concomitant calcium antagonists on treatment efficacy of pravastatin estimated by the decrease in coronary artery diameter after two years of treatment (REGRESS data). Upper graph: pravastatin significantly decreased coronary artery diameter compared with placebo. Lower graph: in the patients receiving no calcium antagonist the benefit of pravastatin was insignificant, while it was highly significant in the patients receiving a concomitant calcium antagonist, suggesting the presence of interaction between calcium antagonists and pravastatin.

the efficacy estimate b_1 was 0.060 (standard error SE=0.016). Calcium antagonists had been given to 60% of the placebo patients, and 59% of the pravastatin patients (χ^2 : $p=0.84$): thus calcium antagonist treatment was not a confounder variable. Also, calcium antagonist medication was not associated with diameter decrease ($p=0.62$). In the patients who did not receive concomitant calcium antagonist medication, the diameter decreases were 0.097 (SE 0.014) and 0.088 (SE 0.014) in patients receiving placebo and pravastatin, respectively ($p=0.71$). In patients who did receive calcium antagonist medication, the diameter decreases were 0.130 (SE 0.014) and 0.035 (SE 0.014), respectively ($p<0.001$). Thus, pravastatin efficacy was, on average, 0.097 to 0.088=0.009 mm in patients without calcium antagonist medication, and 0.130 to 0.035=0.095 in patients with calcium antagonist medication (figure 2, lower graph). The two lines cross, suggesting the presence of interaction between pravastatin and calcium antagonists.

Before statistically testing this suggested interaction it may be wise to assess whether the suggested interaction between the two compounds makes sense clinically. Atherosclerosis is characterised not only by depots of cholesterol but also by depots of calcium in the fatty streaks that consist of foam cells. It does make sense to argue that although they do not reduce plasma

calcium, calcium antagonists may reduce calcium levels in the foam cells and, thus, have a beneficial effect on the process of atherosclerosis, and that interaction with cholesterol-lowering treatment is a possibility.

The following linear regression model was used for testing:⁴

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + e_i,$$

where:

y_i = dependent variable = decrease in coronary artery diameter in the i th patient;

a = intercept;

b_1 , b_2 , and b_3 = partial regression coefficients for the variables (1) treatment modality, (2) calcium antagonist treatment, (3) interaction between (1) and (2).

Let $x_{1i}=1$ denote that patient i received pravastatin ($x_{1i}=0$, if not), let $x_{2i}=1$ denote that patient i received calcium antagonist medication ($x_{2i}=0$, if not), and let $x_{3i} = x_{1i}$ times x_{2i} . The estimates were: $b_3=0.085$ (SE 0.033), $b_2=-0.033$ (SE 0.023), and $b_1=0.009$ (SE 0.026). Notice that b_1 changed dramatically by including the interaction term x_3 in the linear model; this is a general feature of regression models with interaction terms: the corresponding main effects (b_1 and b_2) cannot be interpreted independently of the interaction term. Another consequence is that the efficacy estimate no longer exists, but there are several estimates: in our

Table 4. Results of three treatments for paroxysmal atrial fibrillation in 36 patients in six health centres (two patients per health centre per treatment).

Health centre	Verapamil (No. attacks per patient)	Metoprolol (No. attacks per patient)	Digoxin (No. attacks per patient)	Total (No. attacks per centre)
1	4 6 + 10	10 9 + 19	10 10 + 20	49
2	5 7 + 12	9 11 + 20	11 10 + 21	53
3	4 7 + 11	11 12 + 23	10 13 + 23	57
4	9 10 + 19	6 8 + 14	11 11 + 22	55
5	12 12 + 24	7 7 + 14	12 13 + 22	63
6	11 12 + 23	7 8 + 15	14 13 + 27	65
Total	99	105	138	342

Results are given as numbers of attacks in four weeks.

case there are different efficacy estimates for patients with ($b_1+b_3=0.009+0.085=0.094$) and without calcium antagonist medication ($b_1=0.009$). This difference was statistically significant (interaction test: $p=0.011$).

Analysis of variance for testing interaction

Current cardiovascular trials of new treatments often include patients from multiple health centres, national or international. Differences between centres may affect local results. We might say these data are at risk of interaction between centres and treatment efficacy. As an example: 36 patients with paroxysmal atrial fibrillation were assessed for frequency of attacks while on treatment with three different treatments. Patients were randomly selected in six health centres, six patients per centre, and each patient was given one treatment at random, and so in each centre two patients were given one of the three treatments. Table 4 gives an overview of the results.

The model is: $y=\mu + a + b + ab + e$,
where:

y =dependent variable, estimate for performance of patients;

μ =mean result;

a =fixed effect of the three treatments;

b =random variable associated with health centre;

ab =random interaction effect between treatment and health centre;

e =systematic error.

The computations are (SS=sum of squares)

$$SS \text{ total} = 4^2 + \dots + 13^2 - \frac{(342)^2}{36} = 245$$

$$SS \text{ ab} = \frac{10^2 + \dots + 27^2}{2} - \frac{(342)^2}{36} = 224$$

$$SS \text{ error} = SS \text{ total} - SS \text{ ab} = 245 - 224 = 21$$

$$SS \text{ columns} = \frac{99^2 + 105^2 + 138^2}{12} - \frac{342^2}{36} = 73.5$$

$$SS \text{ rows} = \frac{49 + \dots + 65^2}{6} - \frac{342^2}{36} = 30.67$$

$$SS \text{ interaction} = SS \text{ ab} - SS \text{ rows} - SS \text{ columns} \\ = 224 - 30.67 - 73.5 = 119.8$$

Table 5 gives the ANOVA (analysis of variance) table. The F test for treatment effect is significant. The cumulative numbers of attacks per treatment are, obviously, more different than compatible with random. Also the F test for health centres is significant. The cumulative numbers of attacks per health centre are more different than compatible with random. When studying the data we observe something special and unexpected. The cumulative number of attacks in the health centres 4-6 is 183, while it is only 159 in the centres 1-3. Despite the better cumulative score in health centres 4-6, the treatment 2 seems to perform better in the health centres 1-3. This is unexpected, and may be due to interaction between the efficacy of treatment and the presence of a particular health centre. The F test for interaction confirms that this is so. There is something about the combination of a particular health centre with a particular treatment that accounts for differences in the data. Thus, within the health centres, treatment differences apparently exist, and they tend to be different for different health centres. Perhaps the capacity of a treatment to produce a certain result in a given patient depends on his/her health centre background. Explanations include environmental factors such as social and ethnic factors, investigator factors.

Discussion

Cardiovascular trials do not usually include interaction assessments in the protocol. Results of such assessments are, therefore, post-hoc, and of an exploratory and unconfirmed nature. So, why should they be performed? In cardiovascular research drug-drug interactions and effects of comorbidities on drug efficacies are numerous. It is valuable to account at least post-hoc for such mechanisms. Second, current clinical trials involve heterogeneous health centres, investigators, and patient groups. Accounting for these heterogeneities can be helpful to predict individual responses in future patients, and to develop prediction rules based on the trial data of individuals. Prediction rules such as the Framingham risk score may be developed using trial data to further identify subjects at risk of having a good or bad drug response.

Other reasons for interaction assessments include the following. It may be useful and reassuring to know

Table 5. ANOVA table of analysis for data of table 4.

Source	SS	DF	MS	F	P value
Columns (treatment effect)	73.5	3-1=2	36.75	36.75/11.98=3.07	ns
Rows (centre effect)	30.67	6-1=5	6.13	6.13 /1.17=5.26	<0.01
Interaction(treatment x centre)	119.8	10	11.98	11.98/1.17=10.24	<0.0001
Error (total)	21	18x(2-1)=18	1.17		
Total	245	35			

SS=sum of squares, DF=degrees of freedom, MS=mean square, F=test statistic for F test, ns=not significant.

that in a positive study the benefit in subgroups parallels the benefit in the study overall. For example, in a subgroup analysis of the Dietary Approach to Stop Hypertension (DASH) randomised clinical trial the results were equivalent in different age, gender, and ethnic groups.⁵ Also, in an unexpectedly negative study it may be useful to know if there are certain subgroups that might benefit or be harmed by the treatment. For example, oestrogen/progestin replacement caused cardiovascular benefit in women with high lipoprotein, and harm in those with low lipoprotein.⁶ We should add that the assessment of interaction, otherwise called heterogeneity, is not always wise. In controlled clinical trials a myriad of subgroups can be identified that would qualify for an exploratory examination, and this approach will almost certainly produce one or more spuriously significant interactions. Interaction terms to be assessed should, therefore, make sense clinically. Demonstrating a statistically significant interaction between the treatment effect and the first letters of the patients' Christian names makes no sense, and pursuing such a finding is merely data dredging. We should be alert to the fact that a 'scientific' explanation can be found for any subgroup results in one afternoon PubMed search. As stated by Barrett-Connor in a commentary on the nine positive interactions demonstrated in a subgroup analysis of the otherwise negative Heart and Estrogen/progestin Replacement Study (HERS) trial, biological plausibility is quite easy to theorise, anyone with two hours and a little imagination can do it.⁷

Statistical methods for identifying and integrating interaction terms into the data are limited, and have a limited statistical power. Moreover, if we introduce a large number of interaction terms in a regression analysis, the statistical power to demonstrate statistical significance for the primary endpoint will be reduced. Nonetheless, the assessment of a small number of interaction terms in clinical research can be an important part of the evaluation of new drugs.

The issue of testing interaction is different with meta-analyses of cardiovascular trials.⁸ The aim of a meta-analysis is to obtain a pooled estimate of a treatment effect rather than the study of subgroups. The studies to be included in a meta-analysis are often heterogeneous, and protocols, therefore, routinely

apply a heterogeneity test prior to data pooling. In the presence of a statistically significant heterogeneity, a data pooling may be difficult to accept, and has as an additional problem that confidence intervals are underestimated, because the extra variability between the different trials is ignored.

If a statistically significant interaction is demonstrated post hoc, its existence should be confirmed in a novel prospective clinical trial. If a relevant interaction is expected prior to the trial, its assessment should be properly included in the trial protocol at the planning stage of the trial. Instead of a regression model a factorial trial design is suitable for such purposes.

Linear regression analyses may provide better precision to test interaction than comparison of subgroups.⁴ It is also often more convenient, because it enables all the data to be analysed simultaneously. Different regression models may be adequate for different types of data, e.g., exponential models are more adequate than linear models for risk ratios and mortality data. ■

References

- 1 Riegelman RK. Studying a study and testing a test. Lippincott Williams & Wilkins, Philadelphia, PA, 2005.
- 2 De Craen AJM, Westendorp RGJ. The use of age as a variable in clinical research. *Ned Tijdschr Geneesk* 2005;**149**:2958-63.
- 3 Jukema AJ, Zwinderman AH, et al. for the REGRESS Study Group. Effects of lipid lowering by pravastatin on progression and regression of coronary artery disease in symptomatic men with normal to moderately elevated serum cholesterol levels. The Regression Growth Evaluation Statin Study (REGRESS). *Circulation* 1995;**91**:2528-40.
- 4 Zwinderman AH. Regression analysis for more precision in the data analysis. In: Statistics applied to clinical trials, 3rd edition. Cleophas TJ, Zwinderman AH, Cleophas AF (eds). Springer Dordrecht Netherlands, 2006, pp 141-50.
- 5 Svetkey LP, Simons-Morton D, Vollmer WM, et al. Effects of dietary patterns on blood pressure: subgroup analysis of the Dietary Approaches to Stop-Hypertension (DASH) randomized clinical trial. *Arch Int Med* 1999;**159**:258-93.
- 6 Shlipak MG, Simon JA, Vittinghoff E, et al. Estrogen and progestin, lipoprotein (a), and the risk of recurrent coronary heart disease events after menopause. *JAMA* 2000;**283**:1845-52.
- 7 Barrett-Connor E. Looking for the pony in the Heart and Estrogen/progestin Replacement Study (HERS) data. *Circulation* 2002;**105**:902-3.
- 8 Chalmers I, Altman DG. Systematic reviews. Br Med J Books, Bristol UK, 1996.