ELSEVIER

Infection, Genetics and Evolution 1 (2002) 303–305

Infection, Genetics and Evolution

www.elsevier.com/locate/meegid

Discussion

Inappropriate application of a model for mixed analysis of variance: Some comments on Elena et al. (2001)

David A. Morrison*

Molecular Parasitology Unit, University of Technology Sydney, Westbourne Street, Gore Hill, NSW 2065, Australia

Received 19 December 2001; received in revised form 16 February 2002; accepted 20 March 2002

1. Introduction

Elena et al. (2001) present the results of an in vitro experiment designed to investigate the role of the size of viral inoculum (used to initiate infection) and the mode of transmission (vertical versus horizontal) on adaptation of RNA viruses to their cellular hosts. In the experiment, they used a measure of "relative fitness" of viral cultures to compare the two modes of transmission and two inoculum sizes (small and large) in a series of replicate laboratory flasks, in order to mimic virus evolution. They repeated the experiment four times.

The combined data from these experiments were subjected to a single analysis of variance (ANOVA). The authors describe this analysis as follows: "the linear model used to fit the data was a mixed model II ANOVA where the random factor 'individual host' (flask) was nested within each [experimental] treatment.... The entire experiment was replicated in four independent blocks, allowing the two fixed factors (transmission type and inoculum size) and their interaction to be nested within block." This is thus a four-factor mixed analysis of variance, with two random and two fixed factors, and with three of the factors nested within the other factor (these terms are described in more detail in the subsequent sections).

There are three points that can be made about this linear model and its attendant statistical analysis: (i) the model as specified by Elena et al. violates a basic principle of mixed analysis of variance; (ii) as a consequence, the results of the statistical analysis as presented in Table 1 of Elena et al. do not match the model specified by them; and (iii) a more appropriate model for analysing these data leads to somewhat different conclusions from those presented by Elena et al. These points are discussed in this note.

2. Mixed analysis of variance

An analysis of variance involves a statistical test of the differences between groups of replicate observations defined by a particular experimental treatment—a grouping variable is called a "factor" and the groups are called "levels". The calculations of the analysis produce an *F*-ratio, which is used to decide whether there is a statistically significant difference between the means of two or more of the groups. The *F*-ratio is calculated using the mean-squares, which measure the amount of variability both within and between groups.

When an analysis of variance involves more than one factor, then the calculations used for the *F*-ratios depend on whether the factors are random or fixed and whether they are nested or orthogonal (as described in the following sections). Therefore, Elena et al. have been careful to specify both of these features in the description of the linear model used for their particular analysis, something that is often sadly lacking in biological publications.

For a factor to be considered fixed (or model I), all possible levels of the factor that are of interest for the experimental question must have been sampled in the experiment. For a factor to be considered random (or model II), only a sample of the possible levels of the factor have been included in the experiment. So, Elena et al. treated the factors for transmission mode and for inoculum size as fixed in the model and analysis—the levels of these factors were not arbitrarily chosen, but were deliberately chosen to represent particular repeatable experimental manipulations. The factors experimental block and individual host (i.e. laboratory flask) were treated as random because they were merely a selection of all of the possible flasks or experimental conditions that could have been used in the experiment. These decisions seem to be entirely appropriate for their experimental design.

If factors are orthogonal then every possible combination of the levels of the factors is included in the experiment. For nested factors, different levels of one of the factors occur in combination with only one of the levels of the other

* Tel.: +61-2-9514-4159; fax: +61-2-9514-4003.
E-mail address: david.morrison@uts.edu.au (D.A. Morrison).

PII: S1567-1348(02)00037-0

Table 1
ANOVAs for the experiment described by Elena et al. (2001)

Source of variation	d.f.	M.S.	Analysis of Elena et al.				More appropriate analysis			
			Nesting[a]	*F*-ratio	*F*	*P*	Nesting[b]	*F*-ratio	*F*	*P*
Experimental block	3	11.1768	A	–[c]	0.35	0.795	A	A/D	19.78	<0.001
Transmission mode	4	5.9465	B(A)	B/BC	1.54	0.344	B	B/D	10.52	<0.001
Inoculum size	4	30.2216	C(A)	C/BC	7.80	0.036	C	C/D	53.48	<0.001
Mode × size interaction	4	3.8728	BC(A)	BC/D	6.88	<0.001	BC	BC/D	6.88	<0.001
Individual host	64	0.5651	D(ABC)	D/Res	1.80	0.002	D(ABC)	D/Res	1.80	0.002
Residual	169	0.3145	Res(ABCD)				Res(ABCD)			

[a] All four factors are random.
[b] Transmission mode and inoculum size are fixed factors, while experimental block and individual host are random factors.
[c] There is no exact *F*-test for this factor in this model. Elena et al. calculated a pseudo *F*-test as: A/(B + C–BC).

factors. So, Elena et al. treated the factors for transmission mode and for inoculum size as orthogonal in the model and analysis—the small and large inoculum sizes were tested in combination with both vertical and horizontal transmission, so that all four possible combinations were included in the experiment. The factor individual host (i.e. flask) was treated as nested within the other three factors because each flask was used only once in the experiment, so that the levels of this factor could not appear in combination with more than one level of any of the other factors. This also seems to be entirely appropriate for their experimental design. However, the decision to treat the other three factors as nested within the experimental block factor creates a mismatch between the experimental design and the proposed statistical analysis.

Sokal and Rohlf (1994) note about ANOVAs containing nested factors that: "a crucial point is that groups representing a subordinate level of [a nested] classification must be randomly chosen. . . . The subordinate level of a nested ANOVA is always model II. The highest level of classification in a nested ANOVA may be model I or II." The basic point here is that if the levels of the upper factor in a nested model are random then it is illogical to suggest that the levels of the factors nested within it can be fixed. Alternatively, if any of the nested factors is fixed, then the upper factor cannot be random. This follows logically from the definition of the terms random and fixed—if the levels of the upper factor have been chosen at random then so must be the levels all of the factors nested within it.

The model specified by Elena et al. thus contains a logical contradiction, and any analysis based on it cannot produce meaningful results. This contradiction arises from trying to nest two fixed factors (transmission mode and inoculum size) within a random factor (experimental block).

3. Results of the analysis

Since it is illogical to have fixed factors nested within random ones, statistical analysis computer programs, such as Minitab (Minitab Inc., 2000) and SPSS (SPSS Inc., 2000) automatically set all factors to random if they are nested within a random factor, irrespective of what the user has explicitly specified about those nested factors. This feature is not always notified to the user at the time of the analysis, but the consequences are usually indicated in the output from the program. This feature does not necessarily result in a correct analysis of the data, but it does ensure that the analysis is not internally contradictory.

Elena et al. used SPSS Inc. (2000) to perform the calculations for their analysis. It is clear from the results presented in their Table 1 that their analysis of variance was performed with all four of the factors specified as random, rather than two of them being specified as fixed. This can easily be verified by the *F*-ratios calculated from the mean-squares. Table 1 shows the derivation of the calculations for the *F*-ratios from the analysis presented by Elena et al. The procedures for determining the expected mean-squares, and therefore the appropriate *F*-ratios, for mixed analyses of variance are described by, for example, Winer et al. (1991) and Zar (1999), and are available in computer programs such as that of Dallal (1988).

Thus, the model specified by Elena et al. does not match the actual analysis that they present. Furthermore, the analysis as performed by Elena et al. is incorrect. The factors for transmission mode and for inoculum size *should* be treated as fixed in the model and analysis (as described earlier). Consequently, any analysis that treats these factors as random is inappropriate because it does not match the experimental design. By converting these two factors to random the SPSS program has made the wrong decision when confronted by a logical contradiction in the model being specified for the analysis.

4. More appropriate models

It is straightforward to devise a more appropriate model for the ANOVA needed to analyse the experiment described by Elena et al. Such an analysis is also shown in Table 1. In this analysis, the factors transmission mode and inoculum size are *not* nested within the factor experimental block but

are orthogonal to it. This is the usual way of dealing with experimental repetitions in an analysis of variance, because each of the two levels of the factors transmission mode and inoculum size was included in each experimental repetition so that all possible combinations of levels were included in every experiment. The factor individual host remains as nested within the other three factors. transmission mode and inoculum size remain as fixed factors, while experimental block and individual host are random factors, exactly as in the model originally specified by Elena et al.

Note that this analysis has much more statistical power, as a result of the increased degrees of freedom in the denominator of the *F*-ratio, as well as being more appropriate for the experimental design. Consequently, all of the null hypotheses represented by the different model factors are now rejected by the statistical analysis, if we use the conventional type I error rate of $P = 0.05$. This is quite different from the results of the hypothesis tests using the analysis of Elena et al. For example, they conclude from their analysis: "effect of transmission mode on fitness was found to be not significant", which is now shown to be incorrect. This false conclusion can potentially change their interpretation of the experiment, since testing this hypothesis was one of the prime motivations for the experiment. Furthermore: "we observed homogeneity among the four replicate blocks of our experiment", which is now also shown to be incorrect. It is interesting to speculate about why the experimental repetitions produced different fitness values (on average), although I cannot provide any simple explanation. This important point also raises the issue of the suitable number of "replicate" experiments for such a study—if there is significant variability between the experiments in this experimental system then clearly a small number of repetitions would be inappropriate.

An even more appropriate model for the analysis might also include the interactions between experimental block and both transmission mode and inoculum size. In the analysis described earlier, the mean-squares for these interactions are included in the mean-squares for transmission mode, inoculum size and their interaction. If any of the three possible interactions in this new model is significant, then this would mean that the results of the experiment were quite different among the four experimental repetitions. This is a point that could be important to assess, given the observed statistical significance of the difference in fitness values between the experimental repetitions. The results of such an analysis cannot be presented here in the absence of the original data of Elena et al.

References

Dallal, G.E., 1988. DESIGN: A Supplementary Module for SYSTAT and SYGRAPH, Version 2.0. SYSTAT Inc., Evanston, IL.

Elena, S.F., Sanjuán, R., Bordería, A.V., Turner, P.E., 2001. Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses. Infect. Gen. Evol. 1, 41–48.

Minitab Inc., 2000. Minitab Release 13.1 for Windows. Minitab Inc., State College, PA.

Sokal, R.R., Rohlf, F.J., 1994. Biometry: The Principles and Practice of Statistics in Biological Research, 3rd Edition. Freeman, San Francisco.

SPSS Inc., 2000. SPSS Release 10.0 for Windows. SPSS Inc., Chicago, IL.

Winer, B.J., Brown, D.R., Michels, K.M., 1991. Statistical Principles in Experimental Design, 3rd Edition. McGraw-Hill, New York.

Zar, J.H., 1999. Biostatistical Analysis, 4th Edition. Prentice-Hall, Upper Saddle River, NJ.