ON THE USE OF REGRESSION TECHNIQUES FOR THE ANALYSIS OF SINGLE CASE APHASIC DATA

ANDREW W. ELLIS,* CARMEL LUM† and MATTHEW A. LAMBON RALPH*

*Department of Psychology, University of York, U.K.; †Queen Margaret College, Edinburgh, U.K.

Abstract—Many different properties of objects and their names have been implicated as determinants of aphasic naming accuracy. Many of these variables are intercorrelated, making it very difficult for the experimenter to devise sets of items which differ on one property while being matched on all the others. Faced with this situation, experimenters have been tempted to employ the techniques of regression analysis on aphasic naming data, seeking statistical rather than experimental control over the variables. We report a study in which six aphasic patients attempted to name 139 object pictures on three separate occasions. Multiple regression analyses were carried out on the data for the whole group and the combined data for each patient, and logistic regression analyses on the data from individual administrations of the picture set. The different analyses produce disconcertingly different solutions, casting doubt upon the reliability of the outcomes of these analyses, particularly for single administrations of item sets where the 'significant' predictors only exert small effects. We conclude with a discussion of when, and how, regression analyses might be of use in the analysis of aphasic data. Copyright © 1996 Elsevier Science Ltd

INTRODUCTION

Most aphasic patients have problems with word retrieval and production, and many of them show some degree of consistency from one occasion to another. That is, if an aphasic patient is asked, say, to name a set of pictures of objects on two separate occasions, then the items got right on one occasion will also tend to be got right on the other occasion, and conversely. When such consistency is present, then it is reasonable to ask what attributes or properties of objects or words make them easy or difficult for a given patient. Such information should help to contribute towards understanding what has gone wrong with the process of word retrieval and production in that patient.

Unfortunately, a disconcertingly large number of different attributes and properties of objects and words have been implicated in one study or another as determinants of aphasic naming accuracy [1]. The list includes how familiar an object is in everyday experience [2,3], the frequency with which its name occurs in the language [4,5], the age at which its name is typically learned [2,6,7], the ease with which a mental image of the object can be formed [8], and the number of phonemes in its name [1,9,10].

If an investigator wished to determine the relevance or otherwise of each of these and other factors to the performance of a given patient using traditional factorial methods, it would be necessary to develop sets of items which vary each of the factors in turn while holding all the others constant. That would be a difficult as well as a very time-consuming exercise. Additionally, when the factors of interest are themselves intercorrelated, as they often are (e.g.

Address for correspondence: Andrew Ellis, Department of Psychology, University of York, York YO1 5DD, U.K.

familiar objects tend to be denoted by short, early-acquired, high-frequency names), then it can be hard to find suitable sets of items, and the ones eventually chosen can be 'odd' in various ways [11].

In view of these problems, investigators have been tempted to turn instead to the techniques of regression analysis. Here the patient is asked to respond to a set of items whose familiarity, frequency, length, etc., are known. Success or failure on each item is taken as the dependent variable and a regression analysis is used to discover which, if any, of the predictor variables accounts for a significant independent proportion of the variance [1,2,7].

A recent example of this approach is an investigation of aphasic naming by Nickels and Howard [1]. They report two studies in which regression analyses are used to examine the factors which contribute to determining whether object pictures are named correctly or incorrectly by groups of aphasic patients. In their first study, 12 aphasic patients attempted to name 104 objects on five separate occasions. Responses were scored as correct if the target word, or something which deviated from it by only one phoneme, was produced at any time during the attempt at a picture. Thus, each object could have a score ranging from 0 to 5 for each patient. Two of Nickels and Howard's predictor variables failed to correlate significantly with the results of the group as a whole: written word frequency [12] and a measure of the visual complexity of the pictures. The other six predictor variables—age of acquisition, word familiarity, imageability, concreteness, word length, and operativity (a measure of the discreteness and manipulability of an object [13,14])—all showed significant correlations with overall naming accuracy. However, these predictors all tend to correlate with each other, so it is possible that some of the significant raw correlations are just indirect reflections of the influence of other factors. When Nickels and Howard carried out a simultaneous multiple regression on the group data, only three variables were found to have significant independent predictive power: age of acquisition, operativity and word length (the number of phonemes in the object name).

Nickels and Howard [1] then went on to ask whether the pattern which characterised the group as a whole held for each of the individual members of that group. Separate multiple regression analyses were carried out on the naming data from each patient. Five of the 12 patients showed independent effects of age of acquisition on naming, while seven showed effects of operativity. Only one patient showed an independent effect of word length, despite that variable having been a significant predictor of the performance of the group as a whole. Variables which had not contributed to predicting the group data nevertheless made significant contributions to predicting the results of two individual cases, with one patient showing a frequency effect and another a concreteness effect. None of the patients showed effects of familiarity, imageability or visual complexity.

Study 2 of Nickels and Howard [1; see also 8] involved a new set of 15 patients naming 130 line drawings of objects on one occasion each. A strict scoring criterion was adopted, with only a fully correct initial response being scored as correct. The predictor variables were the same as in Study 1, except that visual complexity was omitted. As in Study 1, written word frequency [12] failed to show even a significant raw correlation with naming accuracy for the group as a whole. Nickels and Howard interpret the lack of a significant frequency effect in either of their studies as implying that previous investigations of aphasic naming that have reported frequency effects may actually have been picking up effects of other, uncontrolled variables which correlate with frequency; for example, age of acquisition and word length. As they note, this echoes studies of word recognition and production in normal subjects which have found that previously reported 'frequency' effects are reduced

or eliminated when these other variables are taken into account [15–18].

When the data for the Study 2 group as a whole were analysed using simultaneous multiple regression, age of acquisition, word length and imageability emerged as significant predictors. As with Study 2, however, analysis of the data from individual cases suggested that what was true for the group as a whole did not necessarily hold for all of the patients within the group. Because each patient only attempted to name each object once, the dependent variable took the form of 0s (incorrect) and 1s (correct). Nickels and Howard argued that the most appropriate form of analysis in such circumstances is a version of regression analysis called ‘discriminant analysis’. This procedure uses the regression equation to classify items and the predictor variables associated with them into groups; in this instance into the group of objects named correctly and the group of objects named incorrectly. Of the three variables that predicted performance for the group as a whole, word length was the most successful predictor of individual results, making a significant independent contribution in 9 of the 15 patients. Age of acquisition was a significant predictor for three patients, and imageability for just one patient. Nickels and Howard argued that the contribution of age of acquisition may have been underestimated because of the inclusion in the analysis of two other variables with which it correlates—frequency and familiarity. Removing frequency had little effect on the outcome as far as age of acquisition was concerned, but with familiarity removed three more patients showed effects of age of acquisition.

In none of the discriminant analyses attempted by Nickels and Howard did any patient show effects of all three of the variables that had predicted performance for the group as a whole. The same was true of the three variables that predicted performance for the group as a whole in Study 1. In contemplating these results, Nickels and Howard drew the by-now standard conclusion that “attempting to characterise aphasic naming generally by examining the performance of groups of aphasics is meaningless as there is no ‘average’ aphasic”, and suggest that, “Rather . . . than examining group performance we would argue that it is more meaningful to observe the range of performance (and patterns of variables which significantly affect performance) across a case series of aphasic patients” [1, p. 1295].

We have no objection to these sentiments in principle: our reservations and concerns have to do with the reliability of the outcome of the statistical procedures employed by Nickels and Howard, and in particular with the trustworthiness of regression analyses of data in which single patients name a set of objects on a single occasion. In their Study 2, Nickels and Howard presented the set of pictures just once to each patient. We have no way of knowing whether a second administration of the same items on another occasion would have produced the same results. For example, just how reliable are the observations that their patients JG and WM show effects of imageability and length on naming accuracy but not effects of age of acquisition, operativity, frequency, familiarity or concreteness when those seven variables are considered together? In this respect it is unfortunate that Nickels and Howard did not analyse the data from the five separate administrations to each patient of the 104 items in Study 1. That would have provided some feel for the replicability of these analyses, and would have allowed for a comparison with the results obtained by analysing the total scores for each patient.

Another concern regarding the Nickels and Howard study is the lack of information about the ‘goodness of fit’ of their analyses. All regression analyses provide statistics which indicate how well the predictor variables as a set are capable of accounting for the data. The contributions of individual variables should only be considered when the entire set has been shown to be capable of predicting the data with a significant degree of success.

In the light of these considerations, the present study set out to investigate the reliability of

regression analyses as applied to the naming performance of individual aphasic patients on different occasions. Six patients named 139 line drawings of objects on three different occasions separated by intervals of at least a week. Regression analyses were carried out on the combined scores on each item for the group as a whole, for the combined scores obtained by each patient across the three administrations of the items, and for each individual administration. This was done using six predictor variables: visual complexity, concept familiarity, imageability, age of acquisition, spoken word frequency, and word length. One of the advances of the present study over previous ones is that the age of acquisition measure was taken from normative data from children rather than adult estimates of age of acquisition. We also used a more up-to-date measure of spoken rather than written word frequency that is based on samples of British English.

METHODS

Subjects

The subjects in this study were six aphasic patients aged between 43 and 73 who had been aphasic for periods of 15 to 78 months before participating in this study. The patients' characteristics, including their aphasia type and severity as determined by the Western Aphasia Battery [19], are shown in Table 1.

Items and variables

One-hundred and thirty-nine black-and-white line drawings of objects were selected. Ninety-eight were taken straight from the set of pictures published by Snodgrass and Vanderwart [19]. Ten were items from Snodgrass and Vanderwart redrawn to improve recognisability for British subjects. The remaining 31 pictures were new items drawn in as similar a style to Snodgrass and Vanderwart as possible. Young adult subjects showed name agreements of 90% or more for these items (i.e. 90% or more gave the designated name as their first response to the item).

Data were obtained for each item on five different factors. Two of the ratings (imageability and rated age of acquisition) were taken from the Oxford Psycholinguistic Database [20]; the remaining ratings were obtained from undergraduate subjects, with at least 20 subjects contributing to each rating [21]. The five measures were:

1. *Concept familiarity.* Subjects were given the written object names and were asked to rate 'the degree to which you come into contact with or think about the concept' on a scale from 1 = unfamiliar to 5 = very familiar.

2. *Imageability.* This scale provides a measure of the ease with which participants can summon up a mental image of an object when given its written name. The scale runs from 1 = with great difficulty to 7 = very easily.

3. *Age of acquisition.* Morrison *et al.* [21] present objective age of acquisition data derived from 280 children aged 2;6 to 10;11 naming 311 black-and-white drawings of objects, including the drawings used in the present study. There were 20 children in each of the 6-month age bands 2;0–2;6 to 7;6–7;11, and a further 20 children in each of the 12-month age bands 8;0–8;11, 9;0–9;11 and 10;0–10;11. The age of acquisition of each object name was taken to be the age of the youngest group in which 75% or more of the children could name that item either unassisted or with the aid of an initial phoneme cue.

4. *Spoken word frequency.* The Celex Lexical Database [22] provides an up-to-date count of spoken word frequency based on samples of British English [23]. The log of this measure was used in the analyses to reduce skew.

5. *Word length.* This was measured as the number of phonemes in the standard British English pronunciation of the name.

Table 1. Case history details of the patients involved in the naming study

	Age (years)	Gender	Aetiology	Months post-onset	Aphasia type	Aphasia severity
Case 1	66	M	CVA	42	Broca	62
Case 2	64	M	CVA	30	Anomic	90
Case 3	43	M	CVA	26	Broca	74
Case 4	73	F	CVA	24	Anomic	64
Case 5	60	M	CVA	78	Broca	63
Case 6	64	F	CVA	15	Anomic	78

CVA: cerebrovascular accident.

Procedure

Each patient was asked to name the 139 black-and-white drawings of objects on three separate occasions, with at least 1 week separating each administration. The order of administration of the items was different on each occasion. No time limit was imposed for a response. The patients' responses were transcribed at the time and also tape recorded for later checking. As in Study 2 of Nickels and Howard [1], a strict scoring criterion was applied, with only a correct initial response being scored as correct.

RESULTS

Table 2 shows the pattern of intercorrelation among the predictor variables and the extent to which each variable correlates with group performance (i.e. the score for each item across the six patients). The only predictor variables to show significant correlations with group performance were age of acquisition and imageability, though familiarity ($P = 0.085$) and word frequency ($P = 0.068$) approached significance. A simultaneous multiple regression on the total scores for each item found a significant overall equation ($F = 3.80$, d.f. = 5,130, $P < 0.01$), showing that the predictor variables taken together had a significant ability to predict performance. However, the only variable to make a significant unique contribution to predicting group performance was age of acquisition ($t = 2.81$, $P < 0.01$). None of the contributions of the other variables approached significance when it came to predicting the performance of the group as a whole.

Simultaneous multiple regression analyses were then conducted on the results for each patient using the total number of correct namings of each item (from 0 to 3) as the dependent variable. This is a similar analysis to that conducted by Nickels and Howard [1] on the data from their Study 1. The results are shown in Table 3. The five predictor variables showed a significant capacity to predict the naming scores combined across the three administrations for Cases 1, 2 and 3. Case 1 showed significant independent effects of age of acquisition, concept familiarity and imageability. Case 2 showed only a significant effect of word frequency. Case 3 showed significant independent effects of age of acquisition and concept familiarity. Cases 4 and 5 showed no significant correlations of the dependent variables with combined performance for the three administrations. Case 6 showed a significant correlation between age of acquisition and naming, and the multiple regression produced a significant value of t associated with that variable, yet the overall regression equation was not significant, suggesting that little trust should be placed in Case 6's age of acquisition effect.

The final set of analyses concern the separate administrations of the picture set. In these

Table 2. Intercorrelation matrix including the correlation of each predictor variable with naming performance for the group as a whole

	FAM	IMAG	AoA	FREQ	PHON
FAM	1.00	-0.021	-0.404	0.514	-0.341
IMAG		1.00	-0.105	0.021	0.078
AoA			1.00	-0.300	0.350
FREQ				1.00	-0.316
PHON					1.00
Naming:					
Correlation: $r =$	0.147 ns	0.175*	-0.271**	0.155 ns	0.070 ns

FAM: rated concept familiarity; IMAG: rated imageability; AoA: normative age of acquisition; FREQ: spoken word frequency; PHON: word length in phonemes; * $P < 0.05$; ** $P < 0.01$; ns: not significant, $P > 0.05$ (one-tailed).

Table 3. Correlations and multiple regression of the effects of the predictor variables on the combined naming scores for individual patients

	PC	<i>F</i> -ratio (d.f. = 5,130)		FAM	IMAG	AoA	FREQ	PHON
Case 1	0.70		Corr:	-0.171*	0.223**	0.242**	0.004 ns	0.019 ns
		5.28**	Regr:	3.131**	1.883 ns	3.939**		
Case 2	0.77		Corr:	0.218**	0.027 ns	0.165*	0.285**	0.185*
		3.26**	Regr:	1.043 ns		0.374 ns	2.027*	0.989 ns
Case 3	0.75		Corr:	0.265**	0.142 ns	-0.289**	0.210*	-0.068 ns
		4.51**	Regr:	2.137*		2.334*	0.636 ns	
Case 4	0.78		Corr:	0.148 ns	0.080 ns	0.144 ns	0.037 ns	-0.074 ns
		1.41 ns	Regr:					
Case 5	0.70		Corr:	-0.037 ns	-0.054 ns	0.008 ns	-0.040 ns	0.074 ns
		0.53 ns	Regr:					
Case 6	0.81		Corr:	0.000 ns	0.125 ns	0.167*	0.069 ns	0.002 ns
		1.11 ns	Regr:			1.96**		

PC: proportion correct across three administrations; FAM: rated concept familiarity; IMAG: rated image-ability; AoA: normative age of acquisition; FREQ: spoken word frequency; PHON: word length in phonemes. Corr: correlation between naming and each variable – value of *r*; Regr: simultaneous multiple regression with all five predictor variables included – value of *t* given for those variables which were significant in the simple correlation; **P* < 0.05; ***P* < 0.01; ns: not significant, *P* > 0.05 (one-tailed).

analyses the dependent variable is either 0 (named incorrectly) or 1 (named correctly). The choice of which multivariate analysis to use was guided by what is known about the performance characteristics of these analyses. Multiple regression is problematic because it is impossible to satisfy the assumption that the errors are normally distributed when the dependent variable has a binary outcome. Nickels and Howard [1] employed discriminant analysis in their Study 2. This analysis can be used to predict a binary outcome (or several outcomes) but it requires that the assumption of multivariate normality be met as well as that the covariance matrices of all the groups be equal. These conditions are difficult to satisfy for the variables employed in these analyses. We decided to use the related technique of logistic regression as it is suited to the prediction of a binary outcome but does not make assumptions about multivariate normality or equality or covariance matrices [24,25].

The results are shown in Table 4. If we focus on the first administration of the pictures, Cases 1, 2 and 3 showed significant model Chi-squares. Case 1 showed a significant independent effect of age of acquisition, Case 2 a significant effect of word frequency, and Case 3 significant effects of concept familiarity and age of acquisition. Cases 4 and 5 showed no significant model Chi-squares and no significant effects of any of the predictor variables. Case 6 showed a significant effect of age of acquisition, but the model Chi-square was not significant.

We might have stopped at that point and made claims about the six cases after the fashion of Nickels and Howard's [1] Study 2, but instead we repeated the administration of the items twice more. The results of the logistic regressions on administrations 2 and 3 turn out to be worryingly different from those of administration 1. Case 1 showed the same age of acquisition effect in administration 3 as he had shown in administration 1, accompanied by a significant model Chi-square, but in administration 2 the model Chi-square was non-significant, as were all the effects of the predictor variables. Case 2, who had showed a frequency effect in administration 1, showed a concept familiarity effect instead in administration 2, accompanied by a significant model Chi-square. In administration 3 the model Chi-square and all the effects of

Table 4. Correlations and logistic regression analysis of the effects of the predictor variables on the naming scores for individual patients on each of the three administrations of the pictures. Values of the Wald statistic are shown for those variables whose raw correlation with naming accuracy is significant

	Ad	PC	Model Chi-square (d.f. = 5)		FAM	IMAG	AoA	FREQ	PHON
Case 1	1	0.72	11.58*	Corr:	-0.047 ns	0.050 ns	-0.250**	0.058 ns	-0.046 ns
				Wald:			8.80**		
	2	0.65	7.87 ns	Corr:	-0.160 ns	0.169*	-0.039 ns	-0.073 ns	-0.003 ns
				Wald:		2.88 ns			
	3	0.73	13.27*	Corr:	-0.104 ns	0.194*	-0.170*	0.012 ns	0.085 ns
				Wald:		2.86 ns	7.03**		
Case 2	1	0.78	11.48*	Corr:	0.135 ns	-0.034 ns	-0.164 ns	0.246**	-0.136 ns
				Wald:				4.64*	
	2	0.74	12.76*	Corr:	0.245**	0.110 ns	-0.168*	0.206*	-0.123 ns
				Wald:	3.87*		0.14 ns	0.57 ns	
	3	0.78	7.28 ns	Corr:	0.099 ns	-0.022 ns	-0.035 ns	0.187*	-0.149 ns
				Wald:				2.36 ns	
Case 3	1	0.73	17.70**	Corr:	0.279**	0.048 ns	0.283**	0.161 ns	-0.084 ns
				Wald:	5.76*		3.92*		
	2	0.74	9.16 ns	Corr:	0.096 ns	0.192*	-0.164 ns	0.171*	-0.049 ns
				Wald:		3.19 ns		1.73 ns	
	3	0.76	7.63 ns	Corr:	0.178*	0.059 ns	-0.162 ns	0.110 ns	-0.002 ns
				Wald:	3.05 ns				
Case 4	1	0.73	2.70 ns	Corr:	-0.018 ns	0.052 ns	-0.059 ns	-0.090 ns	0.022 ns
				Wald:					
	2	0.81	5.19 ns	Corr:	0.138 ns	0.069 ns	-0.0148 ns	0.081 ns	-0.022 ns
				Wald:					
	3	0.76	8.95 ns	Corr:	0.199*	0.052 ns	-0.109 ns	0.092 ns	-0.154 ns
				Wald:	4.45*				
Case 5	1	0.69	3.02 ns	Corr:	-0.024 ns	-0.043 ns	0.021 ns	0.013 ns	0.120 ns
				Wald:					
	2	0.70	4.93 ns	Corr:	0.027 ns	0.102 ns	0.017 ns	-0.097 ns	0.042 ns
				Wald:					
	3	0.70	1.79 ns	Corr:	0.074 ns	0.051 ns	-0.024 ns	0.004 ns	-0.008 ns
				Wald:					
Case 6	1	0.78	7.49 ns	Corr:	-0.038 ns	0.144 ns	-0.171*	0.05 ns	0.009 ns
				Wald:			4.67*		
	2	0.82	1.95 ns	Corr:	0.077 ns	0.024 ns	0.095 ns	0.079 ns	-0.023 ns
				Wald:					
	3	0.82		Corr:	-0.036 ns	0.111 ns	-0.110 ns	0.028 ns	0.019 ns

Ad: administration; PC: proportion correct on each administration; FAM: rated concept familiarity; IMAG: rated imageability; AoA: normative age of acquisition; FREQ: spoken word frequency; PHON: word length in phonemes. Corr: correlation between naming and each variable – value of r ; Wald: value of the Wald statistic in the logistic regression given for those variables which were significant in the simple correlation; * $P < 0.05$; $P < 0.01$; ns: not significant, $P > 0.05$ (one-tailed).

the predictor variables were non-significant. Case 3, who had shown effects of concept familiarity and age of acquisition in administration 1, showed no significant effects at all in either administration 2 or 3. Case 4 turned up a significant effect of concept familiarity in administration 3, but the model Chi-square was non-significant (as it was in administrations 1 and 2 where none of the predictor effects were significant). Case 5 was, at least, consistent, showing no significant effects on any of the administrations, while the tenuous age of acquisition effect of Case 6 in administration 1 disappeared in administrations 2 and 3 where all effects were non-significant.

DISCUSSION

If we compare the results of the logistic regression analyses of the three separate administrations of the picture set and contrast them with the results for each patient's combined scores (Table 3) and the results for the group as a whole (Table 2), then there are several reasons to be cautious. It is clear that these different analyses can produce disconcertingly different solutions.

Analysis of the data for the groups as a whole suggested that group naming performance was affected by age of acquisition and imageability. Although significant raw correlations between naming and imageability were observed for Case 1 in the combined scores and administrations 2 and 3, and in Case 3 administration 2, the independent effect of imageability was not significant in any of the other regression analyses. One problem here is undoubtedly the restricted range of imageability values of picturable items. Case 5, for example, is a deep dyslexic who shows a strong effect of imageability in reading, but not over the range of values shown by the items in the present study.

The age of acquisition effect observed in the analysis of the whole group data recurred rather more often in the data of individual patients. For the combined naming scores, Cases 1, 2, 3 and 6 showed significant raw correlations between naming and age of acquisition, though the effect was only significant in the multiple regression for Cases 1, 3 and 6. Age of acquisition was a significant predictor for administrations 1 and 3 of Case 1, administration 1 of Case 3 and administration 1 of Case 6 (though in the last case the model Chi-square was not significant). Overall, the results would seem to reinforce the importance of age of acquisition as a predictor of aphasic naming [cf. 1,2,3,6,7]. We note that the age of acquisition measure used here is an objective one based on normative data from children not, as has been the case in previous studies, a subjective measure based on adult estimates of the age at which they learned different words. However, we also note that Cases 4 and 5 showed no indication of age of acquisition effects.

At this point we should acknowledge that Case 5 is patient NP, who was reported by Hirsh and Ellis [7] as showing an age of acquisition effect in both naming and reading aloud. That effect was observed by Hirsh and Ellis in two separate regression analyses of NP's spoken naming, one analysis of her written naming, and in her reading aloud of sets of early and late acquired words matched on frequency, length and imageability. One possibility is that NP had changed over the 4 or so years between the time when the data for the Hirsh and Ellis study and the present study were gathered. It is notable that NP's spoken naming was 65% correct in both of the tests in Hirsh and Ellis, whereas in the present study it has risen to an average of 78% correct for very similar items.

Like Nickels and Howard [1], the present study found only weak evidence that word frequency plays a role in determining aphasic naming. The correlation between frequency and group performance was not significant. Cases 2 and 3 showed significant correlations between frequency and combined naming scores, but an independent contribution of frequency in the regression analysis was only found for Case 2. For the individual administrations, frequency correlated significantly with naming for all three administrations to Case 2, but its contribution in the logistic regression was only present for administration 1. Case 3 showed a significant raw correlation between frequency and naming for administration 2, but the effect was again not found in the logistic regression. None of the remaining 14 individual administrations of the items yielded even a significant raw correlation between frequency and naming. The overall impression is that, as Nickels and Howard [1] argued, the role of frequency has been

overestimated in previous studies as a consequence of a failure to control other factors which correlate with frequency, notably age of acquisition.

As for the remaining two predictors, concept familiarity did not correlate significantly with overall group performance. It correlated significantly with the combined naming scores of Cases 1, 2 and 3 but its contribution in the regression analysis was significant only for Cases 1 and 3. Despite having correlated significantly with combined naming for Case 1, it failed to correlate significantly with that patient's accuracy in any of the individual administrations. There were significant correlations between concept familiarity and naming, accompanied by significant Wald values in the logistic regression, for Case 2 administration 2, Case 3 administration 1 and Case 4 administration 3, but in each case those were the only administrations for which concept familiarity emerged from the logistic regression as a significant predictor, and for Case 4 administration 3 the model Chi-square was not significant. Word length did not correlate significantly with the group data. Only Case 2 showed a significant raw correlation between length and combined naming scores, but the length effect was not significant in the regression analysis. Length failed to correlate significantly with accuracy for any of the individual administrations.

What is perhaps most alarming is the lack of consistency between the results of different analyses. The imageability effect found for the group as a whole is not convincingly present for any of the individual patients. Effects found in the analyses of the combined naming scores fail to occur in the analyses of the individual administration (notably the concept familiarity effect for Case 1). Different administrations of the same items only a week or two apart and with similar overall levels of accuracy yield different results in the logistic regression analyses. For example, Case 1's age of acquisition effect in administration 1 disappears almost completely in administration 2, only to reappear in administration 3. Case 2 shows only an age of acquisition effect in administration 1, but in administration 2 that metamorphoses into an effect of concept familiarity. By administration 3, all effects have disappeared. Case 3 shows significant effects of concept familiarity and age of acquisition in administration 1 which fail to replicate in administrations 2 and 3. And so on.

The overall impression is that regression analyses of single administrations of items are too unstable to be relied upon. One reason for this is the phenomenon referred to as multicollinearity. When the predictor variables are intercorrelated, as they inevitably are in this sort of study, then relatively small changes in which items are got right and which are got wrong can dramatically change the outcome of the regression analysis. This is particularly true when the effects under consideration, while statistically significant, are small and account for relatively little of the variance. That is true of both Nickels and Howard [1] and the present study.

Regression techniques are not the simple solution to the problem of intercorrelated variables that they are sometimes presented as. In particular, it is apparent that one analysis of a single administration of a set of items cannot be depended upon to produce stable, reliable results. As a bare minimum, separate analyses of different administrations are needed (ideally with different sets of items). Consideration of whether or not individual variables predict performance should only proceed when the combined set shows a significant capacity to predict success and failure (we note that Nickels and Howard [1] do not report goodness of fit statistics for their analyses of individual patients). Care should be taken to make the distributions of the predictor variables as normal as possible, and investigators should avoid incorporating two predictor variables into an analysis which might be thought to be two measures of the same thing (e.g. imageability and concreteness; frequency and word familiarity) because such pairs of variables can suppress one another. Finally, when a set of candidate variables has been

identified, their influence should be corroborated using sets of items which manipulate those factors factorially. A combination of regression analyses and factorial comparisons should be used to seek converging evidence as to which properties of objects and words determine success or failure in individual cases. We would not want to banish the new-found regression techniques from the armoury of the cognitive neuropsychologist, but as with all statistical methods users need to be aware of their very real shortcomings.

Acknowledgements—This research was supported in part by the Medical Research Council. Matthew Lambon Ralph is in receipt of a studentship from the Science and Engineering Research Council. Parts of this work were presented at the Annual Meeting of British Psychological Society, Warwick, April 1995.

REFERENCES

1. Nickels, L. and Howard, D. Aphasic naming: what matters? *Neuropsychologia* **33**, 1281–1303, 1995.
2. Hirsh, K. W. and Funnell, E. Those old, familiar things: age of acquisition, familiarity and lexical access in progressive aphasia. *Journal of Neurolinguistics* **9**, 23–32, 1995.
3. Funnell, E. and Sheridan, J. Categories of knowledge? Unfamiliar aspects of living and nonliving things. *Cognitive Neuropsychology* **9**, 135–153, 1992.
4. Howes, D. Application of the word frequency concept to aphasia. In *Disorders of Language*, A. V. S. De Reuck and M. O'Connor (Editors). Churchill, London, 1964.
5. Kay, J. and Ellis, A. W. A cognitive neuropsychological case study of anomia. *Brain* **110**, 613–629, 1987.
6. Feyereisen, P., Van Der Borgh, F. and Seron, X. The operativity effect in naming: a reanalysis. *Neuropsychologia* **26**, 401–415, 1988.
7. Hirsh, K. W. and Ellis, A. W. Age of acquisition and aphasia: a case study. *Cognitive Neuropsychology* **11**, 435–458, 1994.
8. Nickels, L. Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word production. *Language and Cognitive Processes* **10**, 13–45, 1995.
9. Ellis, A. W., Miller, D. and Sin, G. Wernicke's aphasia and normal language processing: a case study in cognitive neuropsychology. *Cognition* **15**, 111–144, 1983.
10. Goodglass, H., Kaplan, E., Weintraub, S. and Ackerman, N. The 'tip-of-the-tongue' phenomenon in aphasia. *Cortex* **12**, 145–153, 1976.
11. Cutler, A. Making up materials is a confounded nuisance, or: will we be able to run any psycholinguistic experiments at all in 1990? *Cognition* **10**, 65–70, 1980.
12. Kucera, H. and Francis, W. *Computational Analysis of Present-day American English*. Brown University Press, Providence, RI, 1967.
13. Gardner, H. The contribution of operativity to naming capacity in aphasic patients. *Neuropsychologia* **11**, 213–220, 1973.
14. Gardner, H. Naming of objects and symbols by children and aphasic patients. *Journal of Psycholinguistic Research* **3**, 133–149, 1974.
15. Brown, G. D. A. and Watson, F. L. First in, first out: word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory and Cognition* **15**, 208–216, 1987.
16. Gilhooly, K. J. and Watson, F. L. Word age-of-acquisition effects: a review. *Current Psychological Research* **1**, 269–286, 1981.
17. Morrison, C. M., Ellis, A. W. and Quinlan, P. T. Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory and Cognition* **20**, 705–714, 1992.
18. Morrison, C. M. and Ellis, A. W. The roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition* **21**, 116–133, 1995.
19. Snodgrass, J. G. and Vanderwart, M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory* **6**, 174–215, 1980.
20. Quinlan, P. T. *The MRC Psycholinguistic Database*. Oxford University Press, London, 1992.
21. Morrison, C. M., Chappell, T. D. and Ellis, A. W. An objective measure of word learning age: age of acquisition and other norms for 311 picturable objects. Unpublished paper, Department of Psychology, University of York, 1996.
22. Center for Lexical Information. *The Celex Lexical Database*. Nijmegen, 1993.
23. Sinclair, J. M. *Looking Up: An Account of the COBUILD Project in Lexical Computing*. Collins, London, 1987.
24. Stevens, J. *Applied Multivariate Statistics for the Social Sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.
25. Norusis, M. J. *SPSS Advanced Statistics User's Guide*. SPSS, Chicago, IL, 1990.