Logistic regression for risk factor modelling in stuttering research

Phil Reed a, *, Yaqiong Wu b

a Swansea University, UK
b University College London, UK

ABSTRACT

Objectives: To outline the uses of logistic regression and other statistical methods for risk factor analysis in the context of research on stuttering.

Design: The principles underlying the application of a logistic regression are illustrated, and the types of questions to which such a technique has been applied in the stuttering field are outlined. The assumptions and limitations of the technique are discussed with respect to existing stuttering research, and with respect to formulating appropriate research strategies to accommodate these considerations. Finally, some alternatives to the approach are briefly discussed.

Results: The way the statistical procedures are employed are demonstrated with some hypothetical data.

Conclusion: Research into several practical issues concerning stuttering could benefit if risk factor modelling were used. Important examples are early diagnosis, prognosis (whether a child will recover or persist) and assessment of treatment outcome.

Educational objectives: After reading this article you will: (a) Summarize the situations in which logistic regression can be applied to a range of issues about stuttering; (b) Follow the steps in performing a logistic regression analysis; (c) Describe the assumptions of the logistic regression technique and the precautions that need to be checked when it is employed; (d) Be able to summarize its advantages over other techniques like estimation of group differences and simple regression.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

The prediction and control of behaviour is the stated aim of behavioural science. Perhaps less grandiosely, this aim could be restated as a quest to identify the variables that have important effects on subsequent behaviours. In whatever way this aim is actually stated, it is clearly central both to the development of an effective clinical science and to the development of successful individual intervention programmes. Indeed, the relative success of behavioural science in turning its laboratory findings into applications in the clinical field may owe much to its focus on establishing functional relationships (i.e., predictive relationships) between variables. In contrast, cognitive science, with its tendency to focus on uncovering and understanding the structural factors (i.e., the underlying mechanisms responsible for generating behaviours), be they mental, cognitive, or neurological, in nature, has had less noticeable success in the clinical arena.

If it is accepted that identifying the functional relationships between variables, those relationships that indicate which predictor variables can potentially be manipulated in order to produce changes in outcome variables, is critical to enhancing
clinical practice, then establishing appropriate research methods to uncover and analyse these predictive associations is critical to forwarding the field. Knowing the kinds of variables that typically predict the occurrence of subsequent behaviours has multiple advantages for the applied psychologist. For example, the knowledge that self-stimulatory behaviour, or self-harming behaviour, is typically a function of only a limited number of kinds of reinforcing- and setting-events, has helped in the control and remediation of these problems (Carr, 1977; Iwata et al., 1994). Similarly, identifying the psychological characteristics of individuals that predict who is likely to benefit from a particular medical intervention has allowed treatment to be better directed (e.g., Jenkins, Stanton, & Jono, 1994) and resources better managed (e.g., Rutledge, Adler, & Friedman, 2011).

Clinically, at the level of the individual patient, such relationships between variables and behaviours are established through a range of functional analytic techniques. Although they are of critical benefit to the individual patient, these functional analyses can be highly idiosyncratic, and the relationships uncovered may only relate to one individual. Thus, they often lack the generality to initially help to focus a search for the critical variables for a new patient — that is, functional analyses tend to identify specific factors that operate for an individual, and not classes of events that may apply across larger populations of individuals.

Given this, researchers have attempted to develop a range of analytic statistical tools that can take large-sample data and produce algorithms that relate sets of variables to subsequent behaviours in order to establish predictive relationships that can help guide and shape interventions to control the targeted behaviours. Such statistical tools include multiple-regression, Kaplan–Meier plots, and logistic regression techniques. Each of these approaches has its own particular uses and advantages. The current article focuses on the use logistic regression techniques in the context of research on stuttering that is an increasingly used and important tool in many practical areas concerning clinical practice in general (see Kawamoto, Houlihan, Balas, & Lobach, 2005).

2. Logistic regression

Logistic regression is used to analyse the relationship between a single predictor, or several predictors, and an outcome that is dichotomous in nature (such as the presence or absence of an event). This form of regression analysis has become an increasingly employed statistical tool, especially over the last two decades (see Oommen, Baise, & Vogel, 2011, for review), although its original can be dated back to the nineteenth century (see Cramer, 2002). It is widely regarded as the statistic of choice for situations in which the occurrence of a binary (dichotomous) outcome is to be predicted (see Hosmer & Lemeshow, 2000; King & Zeng, 2001; but see Tu, 1996, for discussion of some alternative techniques).

In addition to its many uses for developing models that will predict events in the physical sciences (e.g., Lopez & Sanchez, 2009), economics (Boyacioglu, Kara, & Baykan, 2009; Karp, 2009), political sciences (King, Tomz, & Wittenberg, 2000), and in medicine (see Fleck et al., 2005; Jiang, El-Kareh, & Ohno-Machado, 2011), logistic regression has an increasing use in medical and psychological contexts. Examples of the use of logistic regression in the latter settings include study of the factors that predict whether an improvement or no improvement will occur after an intervention (e.g., Fleck et al., 2005; Khan et al., in press), or the presence or absence of stuttering in relation to a variety of factors (e.g., Howell & Davis, 2011; Reilly et al., 2009).

In fact, the prediction of dichotomous outcome variables are especially useful in clinical research settings, where knowing the factors that predict whether or not somebody will show an improvement due to treatment, or whether they will be likely to experience a clinically-significant event (e.g., stutter, stroke, etc.), is vital to decisions about the course of the treatment. In such cases, clearly specified dichotomous outcomes are of prime importance, and, indeed, they may be the only outcome measures that are routinely and consistently taken in clinical practice.

As logistic regression has such widespread uses, there are many very good introductory statistical texts that give very clear explanations of the technique (e.g., Agresti, 2007; Burns, Burns, & Burns, 2008; Hosmer & Lemeshow, 2000; Howell, 1997). Moreover, the steps necessary to conduct such an analysis using statistical software packages like SPSS are also clearly described in many texts (e.g., Burns et al., 2008; Muijs, 2010), on many websites (e.g., Wuensch, 2009), and even on YouTube videos. Moreover, specific introductions to the use of logistic regression in particular contexts have been provided by Greenhouse, Bromberg, and Fromm (1995) for recovery after stroke, and by Peng, Lee, and Ingersoll (2002) for its use in educational settings.

However, logistic regression has also began to be used with an increasing frequency in investigations into various aspects of stuttering (e.g., Dworzynski, Remington, Rijstdijk, Howell, & Plomin, 2007; Howell & Davis, 2011; Jones, Onslow, Harrison, & Packman, 2000; Reilly et al., 2009). In cases where logistic regression has been applied to the analysis of stuttering behaviour, a question is posed in which the factors associated with a dichotomous outcome are investigated; such as, which factors are associated with the onset of stuttering (e.g., Reilly et al., 2009), or the persistence of stuttering (e.g., Howell & Davis, 2011), or with treatment–induced recovery from stuttering (e.g., Jones et al., 2000)? In all of these cases, the outcome is a variable that can take the form only of a ‘yes’ (e.g., recovered) or a ‘no’ (e.g., not recovered). For example, one question of some importance in this area concerns identifying the factors that are associated with recovery from stuttering, which can occur spontaneously (i.e., in the absence of treatment) for a large number of individuals (Yairi & Ambrose, 2004). Knowing which individuals are likely to recover if left alone is an important question to answer, both in terms of efficient allocation of resources (Howell & Davis, 2011), but also as it has been argued for some disorders that unnecessary treatment may actually be harmful (see Cook & Rustin, 1997). In these cases a definition of recovery was produced so that participants could be
coded as having recovered or not, and this binary outcome variable was then related to a variety of predictors (see Howell & Davis, 2011).

2.1. A simple example of logistic regression

Given the potentially wide ranging usefulness of the logistic regression technique for stuttering research, an example of a simple use of the procedure may serve to highlight the mechanics and interpretation of the analysis. To do this, some mock data were generated, for 220 participants, regarding the impact of the number of hours of treatment per week (the predictor) and whether there was a recovery from stuttering (the dichotomous outcome). While these are mock data, the amount of intervention necessary to affect a recovery is question of some practical significance, and is a question which has been posed in various contexts, including stuttering (see Jones et al., 2000; Osborne, McHugh, Saunders, & Reed, 2008).

Fig. 1 shows a scatter-plot of these mock data, relating recovery to hours of treatment per week. Although not particularly interesting in itself, this figure serves to highlight a number of important issues regarding why logistic regression is to be preferred in these contexts over other regression procedures. There are a number of things to note immediately based on an inspection of these data. The first is that the outcome chosen for analysis in this example can only take one of two values (0 = not recovered and 1 = recovered), any regression technique that has the possibility of predicting any other value is clearly inappropriate for such data (it should be noted that it is not always the case that the outcome has to be divided into two categories, sometimes the outcome will be a continuous variable; whether to actually divide this into two will depend on the design chosen) secondly, the relationship between the predictor (hours per week) and outcome (recovery) cannot be termed linear, but are best described by an S-shaped (‘sigmoidal’) curve; and thirdly, the variance in the outcomes (recovery) is much smaller at the extreme values of the predictor (intervention time per week) than it is at the central values. This tendency can be seen more easily in the plot of these values displayed in Fig. 2. This figure represents the mean recovery rate at each level of treatment intensity (not a particularly appropriate statistic), but more importantly, it shows the confidence intervals around those means. Inspection of these confidence intervals reveals much larger intervals (variance) in the middle values of intervention time per week than at the extreme values. These features, especially the latter concerning unequal variance in the outcome variable across all values of the predictor variable, make such data typically unsuitable for simple regression analyses (see Howell, 1997, and section on alternative techniques below).

As these outcome data are not linearly related to the predictor, this means that there are a number of stages that are required in a logistic regression to transform the data so that they do show a linear relationship between predictor and outcome. These stages are: (1) calculating the probabilities of outcomes at each value of the predictor; (2) converting this probability to an odds ratio; and (3) taking the natural logarithm of the odds ratio. Taking an example from the above data displayed in Fig. 1. Firstly, for each of the values of the predictor (intervention hours/week), the probability of an outcome (recovery from stuttering) is calculated. For example, if, for an intervention time of 15 h a week, there were 10 cases of recovery, and 4 cases of continued stuttering, then the probability of recovery for that intervention time is: 10/14 = 0.714. Secondly, the odds ratio for recovery (i.e., the probability of recovery divided by the probability of stuttering) is calculated; so, for the same level of intervention time per week employed in the above example (i.e., 15 h/wk), for which the probability of stuttering was 0.714, the probability of not stuttering would be 0.286 (i.e., 1 − 0.714), and so the odds ratio of recovery is: 0.714/0.286 = 2.497. Finally, to make the data linear, the natural logarithm (i.e., using a base 2.71828) is taken of the odds ratio (which produces a value of 0.915). The log will be positive for odds greater than 1, and negative for odds less than 1. This final stage is called the ‘logit transformation’.

The data are now in a form in which there is a linear, rather than a sigmodal, relationship between the predictor variable (i.e., hours of intervention per week) and the outcome (i.e., recovery of stuttering), and the questions for the logistic analysis become: (1) what is the best way to describe this relationship between predictor and outcome, and (2) is this relationship
statistically significant – does knowing something about the predictor variable enable a better prediction of the outcome than knowing nothing about the predictor? The logistic regression solution regarding the relationship between the predictor (intervention hours per week) and the outcome (recovery) is not unlike those that are noted for standard regression equations, and the logistic regression equation takes the form:

$$\log\text{odds} = \log\left\{\frac{p}{1-p}\right\} = b_0 + b_1 \times \text{(predictor value)}$$

where $b_1$ is the amount of increase in the log odds of the outcome given by an increase in one unit of the predictor, and $b_0$ is the intercept of the model. In the example, the chances of recovery based on the number of hours of intervention are given by:

$$\log\text{odds of recovery} = -8.0 + 0.32 \times \text{hours per week}$$

In this case, the coefficient for the predictor is 0.32, which means that a one hour increase in the intervention time per week would increase the log odds of stuttering recovery by 0.32. As noted by Howell (1977, p. 554), expressed this way, this is not a particularly useful thing to know. In order to make this result of more practical significance, this value (0.32) has to be exponentiated, which gives a value of 1.38. The result means that a one hour increase in the intervention time per week multiplies the odds of recovery from stuttering by a factor 1.38.

This is a useful result in practical terms, of course, but it also has to be assessed for its statistical significance (this would be done at the same time as generating the above logistic model, and the above calculation would only be conducted if the regression proved to be statistically significant). To determine the statistical significance of the relationship between the predictor and outcome, logistic regression uses a maximum likelihood method, which discovers the precise form of the equation that maximizes the chances of predicting the outcome based on the predictor (King & Zeng, 2001). The degree to which the resulting logistic regression equation allows accurate prediction of the outcome is then compared to the case in which no predictors are known (i.e., whether the logistic regression equation performs better than a random prediction).

To achieve this end, the likelihood of observing the outcomes that were actually obtained based on the logistic model, and also based on no predictors, are calculated. The result of this process is often a small number, and to enhance its usability, twice the natural logarithm of this number is used – producing the 2 log likelihood (2LL) value; and this value is the basis for the test of significance. As probabilities are always less than one, the logs of these numbers are always negative, and using a negative value (−2LL) turns them positive. The test is then to compare the difference between the −2LL value for the logistic regression, and the −2LL value for the no predictor model, which is done using a chi squared statistic. A perfect fit between the model and the data would give a −2LL value of 0, and, as the deviation from the predicted fit and the actual data increases, the chi square value increases. (An alternative test for model fit is to compare an estimated model with a so-called "saturated model", which has as many parameters as observations, and would always obtain perfect fit.)

The results from an SPSS analysis (SPSS 16.0) presented in Fig. 1 are shown in Table 1. Here the initial results (Block 0) describe the fit of the model when there are no predictors. In the first table it shows that the model is attempting to predict recovery (the outcome coded as 1), and that the probability of the results of the chi square predicting this (arbitrarily
Table 1
SPSS output from logistic regression conducted on the example data displayed in Fig. 1.

**Block 0: beginning block**
Classification table

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recovered</td>
<td>No recovery</td>
<td>Recovery</td>
<td>Percentage correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>110</td>
<td>.0</td>
</tr>
<tr>
<td>Step 0</td>
<td>Recovered</td>
<td>0</td>
<td>110</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>Recovery</td>
<td>0</td>
<td>110</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>Overall percentage</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variables in the equation

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 0</td>
<td>Constant</td>
<td>.000</td>
<td>.135</td>
<td>.000</td>
<td>1</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Block 1: method = enter**
Omnibus tests of model coefficients

<table>
<thead>
<tr>
<th></th>
<th>Chi-square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1: method = enter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td>Step</td>
<td>155.890</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Block</td>
<td>155.890</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>155.890</td>
<td>1</td>
</tr>
</tbody>
</table>

Model summary

<table>
<thead>
<tr>
<th></th>
<th>–2 Log likelihood</th>
<th>Cox &amp; Snell R square</th>
<th>Nagelkerke R square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>149.095</td>
<td>.508</td>
<td>.677</td>
</tr>
</tbody>
</table>

Variables in the equation

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>time</td>
<td>.321</td>
<td>.042</td>
<td>57.162</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>–8.001</td>
<td>1.066</td>
<td>56.308</td>
<td>1</td>
<td>.000</td>
</tr>
</tbody>
</table>

a Constant is included in the model.

b The cut value is .500
c Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

selected) outcome is no greater than chance. The lower results (Block 1) refer to the fit of the model when intervention time per week is added as predictor, showing the significance of the model (labelled ‘Omnibus Test’), and the –2LL value (labelled ‘Model Summary’). This shows that there was a statistically significant improvement in the goodness of fit as a result of the logistic model; meaning that knowing about the value of the predictor variable allows a better prediction of the outcome than not knowing about the value of the predictor. The coefficients in the logistic equation are also shown in the final table displayed (labelled ‘Variables in the Equation’). Here the beta values (B = 0.321) of the predictor time are shown, as are the B values for the constant (–8.001), which form the values in the above equation. Of course, there are a great many statistical packages, and each will have its own particular printout of the results, but these are the key features to search for in those printouts.

There are a number of alternative methods of analysing the results instead if using the −2LL statistics, such as the Wald statistic, but most analysis of these statistics suggest that the −2LL statistic is preferable (see Darlington, 1990; Hosmer & Lemeshow, 2000; see Johnson, 1986, for further discussion).

2.2. Additional uses of logistic regression

The above example is a straightforward description concerned with determining the extent to which one predictor (i.e., intervention time per week) accurately classifies outcomes (i.e., recovery from stuttering). In the process it also provides an estimation of the degree to which the predictor impacts on the outcome (that is, the odds ratio is also an effect size statistic). However, in many cases, this is not the only question that research seeks to ask by using a logistic regression. There are at least three further uses to which logistic regression has been put in the context of exploring predictors of stuttering outcomes: (1) to explore the effects of, and relationships between, multiple predictors (e.g., Jones et al., 2000; Reilly et al., 2009); (2) to determine which of a range of potential predictors actually are important (e.g., Holland, Greenhouse, Fromm, & Swindell, 1989; Howell & Davis, 2011); and (3) to explore whether newly explored variables add to the predictive validity of already established models (e.g., Fleck et al., 2005).
If more than one predictor variable is to be studied, the procedure outlined above is conducted using all of the predictor variables together. This will produce a logistic regression equation of the same form as in the simple case, but with more predictors. The significance of the \(-2\text{LL}\) for each predictor refers to the independent contribution of that predictor variable to the outcome (as in the case of a multiple regression). An example of this use of logistic regression in a stuttering context was reported by Jones et al. (2000), who used the percentage of words stuttered at the start of the intervention (the Lidcombe programme), and also the time between the child’s stuttering onset and treatment initiation, as predictor variables and continued stuttering or not as the outcome. Both predictors were revealed as significant in the logistic regression, suggesting both had an independent impact on stuttering recovery.

Another goal of research that might employ logistic regression is to determine which, out of a set of potential predictors of an outcome, are the most important, and which can be discarded. This is often a question asked during the exploratory phases of a research project, when little is known about the predictors of the outcome. A study concerning the early predictors (at a younger age) of later stuttering (at an older age) that was reported by Howell and Davis (2011), illustrates this approach (see Holland et al., 1989, for a slightly different method). In the report by Howell and Davis (2011), seven predictors of stuttering were entered into a logistic regression, using the ‘backward stepwise’ method, to establish the initial model (i.e., with all predictors).

Backward stepwise regression is the preferred method of exploratory analyses for logistic regression in comparison to forward regression – starting with the ‘best’ predictor and adding new ones until no increases in predictive power are noted. This is thought to be because the rule for including or excluding predictors involves significance testing of the effects of including or excluding variables. With smaller samples, a traditional level of significance, such as \(p < 0.05\), is too conservative, given too little power, to identify important predictors (Steyerberg, Eijkemans, Harrell, & Habbema, 2000). In backwards stepwise regression, the analysis begins with a model comprising all of the selected putative predictor variables, and predictor variables are eliminated from this model through an iterative process. The predictor that had the least impact on how well the model fitted the data (i.e., the one with the lowest \(-2\text{LL}\) value) is removed initially, and the impact on the overall model’s predictive ability of removing the weakest predictor is studied. The principle of whether a predictor was removed from the model was whether the \(-2\text{LL}\) value of the model decreased significantly after the removal of the predictor. If a predictor was removed, and the \(-2\text{LL}\) was significantly reduced, then that predictor should be kept. This process continues until all of the predictors that could be removed without affecting the statistical significance of the overall model in predicting the outcome had been removed. The results of this analysis reported by Howell and Davis (2011) showed that all variables could be removed, except that of stuttering severity at an earlier age; thus, stuttering severity was the only significant predictor that indicates continuation of stuttering at a later time. In this example, the final logistic regression equation for the log odds of recovery was: \(7.705 + 0.277 \times \text{SSI-3 score}\), which meant that the risk of stuttering goes up by 0.277 for every point increase in SSI-3 score.

The last approach that is commonly used in medical contexts is to determine if a new set of predictors improve on an existing model when they are added. This question is often asked in more well-developed areas, where some of the predictors of an outcome are already established. In the example above reported by Howell and Davis (2011), the baseline model was simply all of the measured predictors, and this was refined (honed down) during the course of the analysis. However, it may also be the case that specific models are the object of the analysis; that is, the contribution of hypothesised predictors over and above already known predictors is sought. For example, in a different context, Fleck et al. (2005) wanted to explore the factors that are associated with intensity of depression. To do so, they used a hierarchical logistic regression, starting with the more basic (well known) predictive factors in the model, such as socio-demographic data, length of depressive disorder, and co-morbidity, and determining the extent to these impacted the outcome. After the goodness of fit of these predictors was established, they then assessed whether adding factors from sets of variables, such as health-related quality of life, key life-events, and place of treatment, improved the predictive accuracy of the model.

2.3. Assumptions and limitations for logistic regression

As with any statistical technique, there are a number of assumptions that underlie the use of logistic regression, and a number of limitations regarding the circumstances under which it can be employed. In terms of the assumptions underlying the use of a logistic regression, these are fairly simple and straightforward, and are set out in Table 2 (see Burns et al.,

| Table 2 |
| Assumptions of logistic regression. |
| **Logistic regression does not assume a linear relationship between the dependent and independent variables (but between the logit of the outcome and the predictor values)** |
| **The dependent variable must be a dichotomy (2 categories)** |
| **The independent variables need not be interval, nor normally distributed, nor linearly related, nor of equal variance within each group** |
| **The categories (groups) must be mutually exclusive and exhaustive; a case can only be in one group and every case must be a member of one of the groups** |
| **Larger samples are needed than for linear regression because maximum likelihood coefficients are large sample estimates** |

Adapted from Burns et al. (2008).
2008). Inspection of these assumptions shows that this technique can be employed somewhat more flexibly than traditional regression techniques, making it suitable for many clinical relevant situations.

However, in addition to the basic assumptions required for use, there are a number of considerations that need to be taken into account that limit the types of data sets to which this regression analysis can be applied. Discussion of these issues, in the context of stuttering research, will illuminate these limiting factors, allow better deployment of this procedure, and, importantly, allow better interpretation of the strength of research that uses this technique.

2.3.1. Sample size

There are a number of ways in which the sample size can impact on the validity of the conclusions drawn from studies employing logistic regression. Of course, there are the usual provisos about extrapolating from small, and potentially unrepresentative, samples. These considerations apply to almost all statistical analyses. However, there are also issues that are specific to logistic regression regarding the potentially detrimental impact of small samples on the results. These sample size issues require some discussion, as they can effect the calculation of the logistic regression.

Of particular concern in the context of logistic regression is the relationship between the sample size and the calculation of the odds ratio (see Clayton & Hills, 1993; Greenland, Schwartzbaum, & Finkle, 2000; Jewell, 1984). The size of the odds ratio (e.g., the probability of recovery divided by the probability of no recovery) that is derived from a logistic regression is inversely related to the size of the sample, and this is often referred to as the ‘sparse-data’ problem (see Greenland et al., 2000). As the sample size decreases, the bias in the odds ratio produced by that sample away from the actual odds ratio becomes larger, and, thus, it gives a poor estimate of the actual population effect (see Clayton & Hills, 1993). In fact, in the worst case scenario, it has been suggested that the bias in the odds ratio due to sparse data sets can double the odds ratio estimates (Greenland et al., 2000), leading to unimportant variables being treated as statistically significant (see Lukacs, Burnham, & Anderson, 2010).

Nemes, Jonasson, Genell, and Steineck (2009) have produced estimates of the bias introduced into the calculation of odds ratios by various sample sizes. They have concluded that odds ratios are overestimated in small samples due to the inherent properties of logistic regression models. A representation of the results from their simulations is shown in Fig. 3, which reveals that the mean deviation from the actual population odds ratio (the bias) becomes smaller with larger sample size. Moreover, the range of odds ratio values that are obtained becomes much more constrained as the sample size increases (noted by the contracted inter-quartile range of odds ratio estimates produced by repeated sampling of a population). Their calculations also suggest that this bias effect is also made worse if the sampling distribution is skewed; large samples protect from this added problem.

This ‘sparse-data’ problem is particularly acute when the research question requires examination of large numbers of covariates, or involves comparison between matched sets of individuals (see Jewell, 1984). Examples of this form of research are when the question involves comparison across a number of treatment sites (e.g., Fleck et al., 2005), or comparison between groups, such as twins (e.g. Dworzynski et al., 2007). Although it should be noted that neither of these specific pieces of research noted above fall foul of this sparse-data problem themselves, they are examples of sub-dividing a sample, which may well produce sparse-data issues. In this context, it might also be noted, briefly, that the ‘conditional logistic
regression’ procedure, which was developed putatively to avoid the ‘sparse-data’ biases that can arise in ordinary logistic regression analysis (e.g., Clayton & Hills, 1993), may also be susceptible itself to these problems (see Greenland et al., 2000).

To counteract these problems, large samples should be used wherever possible with logistic regression (see Burns et al., 2008; Weiss, 2004). The number of subjects that has been suggested to be appropriate in order to avoid the worst effects of sparse-data bias varies from author to author: for instance, King and Zeng (2001) fixed the minimum sample size to avoid the problem at around 200 participants; whereas Nemes et al. (2009) suggested a figure closer to 500 participants as being needed.

In terms of the implications of this discussion for research on stuttering, some studies clearly surpass these minimal suggested numbers of participants (e.g., Dworzynski et al., 2007; N = 12,892; Reilly et al., 2009; N = 1619), some reports are on the borderline for these numbers (e.g., Jones et al., 2000; N = 250), and some are clearly below the minimum numbers suggested (e.g., Howell & Davis, 2011; N = 132; Kingston, Huber, Onslow, Jones, & Packman, 2003; N = 66). That is not to say that the results from these studies are to be ignored, but, rather, these results may have over-estimated the degree to which the factors chosen for study predict the outcome.

An additional, and related, issue concerning the impact of sample size on logistic regression concerns the outcomes of meta-analyses conducted on the findings derived from logistic-regression studies (see Chang, Lipsitz, & Waternaux, 2000). It is commonly thought that combining studies will enhance the power of the analysis, and enable relationships to be noted that may be obscured in smaller samples (see Lipsey & Wilson, 2001). However, this is not necessarily the case with logistic regression studies (or, indeed, with studies using any technique) where the sample sizes of the individual studies are small to moderate. Combining the results of such studies, given the relationship between the sample size and the over-estimation of the odds ratio, will produce larger than actual relationships between predictors and events (see Nemes et al., 2009). Under these circumstances, rather than conducting a meta-analysis on the results of the separate studies, the data need to be combined into one larger sample, and the logistic regression re-run – this approach obviously assumes compatibility of the data (see Kingston et al., 2003, for an example), which may not necessarily apply if the data from the Lidcombe treatments that were obtained in different countries are pooled (see Howell & van Borsel, 2011).

2.3.2. Sampling rare events

In some cases, research involves exploring the predictors of events that occur in the population with unequal frequencies to one another; such as the presence or not of a catastrophic event (see King & Zeng, 2001), or a rare medical condition (see Jiang et al., 2011). In the current context relating to stuttering research, such uneven outcomes could be the presence versus absence of current stuttering (estimated to be about 1:100; e.g., Craig, Hancock, Tran, & Craig, 2002), or lifetime stuttering (estimated to be about 1:20; see Mansson, 2000), or it could be recovery or persistence of stuttering (suggested to be around 4:1; e.g., Yairi & Ambrose, 2004). Such imbalanced outcome data are becoming increasingly recognized as an issue, especially as large datasets are routinely collected and mined (e.g., Ford et al., 2012), and the problems that this produces for logistic regression, and their potential solutions, have been widely discussed (see Weiss, 2004, for a review).

In particular, when a population has such an imbalance in the probability of outcomes, it introduces two related problems when using logistic analysis; (a) the ‘class imbalance’ itself – that is, there is more of one type of event than the other; and (b) the ‘sampling bias’, where the study includes more or less of one type of outcome event than actually occur in the population (see Oomen et al., 2011). A not uncommon example in the field of stuttering research would be if the population has an outcome ratio of 20:1 between the classes of events (such as for lifetime stuttering prevalence). In itself, this represents a class imbalance, but in sampling this population, the sample may further underestimate the probability of the rarer event, which is a hard event to sample (see Howell, 1997; Weiss, 2004), and it may produce a sample ratio of, say, 40:1 (this is sampling bias). Both class imbalance and sampling bias have implications for the results and conduct of studies employing logistic regression, and both of these biases interact with the size and type of the sample studied.

Weiss (2004) has noted that when there are rare events (either in absolute or relative terms), then sampling these data will produce fewer of the rare cases than might be expected. This will lead to a reduced variation in the outcomes represented in the sample, and this will also make it harder to assess the relationship between the predictors and the few instances of the rare event (see also King & Zeng, 2001). That is, the presence of rare, and potentially underestimated events, will seriously impact the power of the logistic regression. Additionally, in logistic regression employing the –2LLR method of statistical significance testing, discussed above, the under-prediction of the actual probabilities of events is proportional to the sampling bias (Oomen et al., 2011; see Howell & Davis, 2011, for a discussion related to stuttering research).

Such concerns about data sets that contain very few events in one category have prompted researchers to collect large numbers of cases in order to capture enough of the rarer outcomes (see Dworzynski et al., 2007; Reilly et al., 2009, for examples of large N stuttering research). Unfortunately, this is not only a time consuming, and costly, research strategy, but it can also, under some circumstances, lead to the collection of data on inappropriate participants – that is, on participants unlikely ever to be subject to the event in question, thus, undermining the representativeness of the sample (see King & Zeng, 2001, for a discussion). This is not to say that such large-scale sampling strategies are always inappropriate if the targeted group comprises individuals who potentially could be subject to both outcomes (see Howell, 2012, in this volume). However, inclusion of participants in the sample who would be highly unlikely to display both outcomes should be avoided (e.g., the inclusion of the country Switzerland in a study looking at the predictors of countries engaging in war).

There are a number of alternative sampling strategies that have been developed in order to overcome this sample bias problem for logistic regression (and related techniques). These strategies have been outlined by Kamei, Monden, Matsumoto,
Kakimoto, and Matsumoto (2007; see also Table 3). Many of these procedures are useful for data mining techniques (Weiss, 2004), but two of the most commonly employed procedures are termed ‘over-sampling’ (i.e., over-representing the rare events observed in the sample), and ‘under-sampling’ (i.e., under-representing the number of more common events obtained in the sample). Both of these techniques attempt, in different ways, to equate the numbers of cases of each outcome. In particular, the under-sampling approach to rectifying sample bias has often been used in epidemiological research, where it is referred to as a case–control design (Breslow, 1996). In this procedure, researchers include all of the rarer events that they have sampled, and employ a random proportion of the more common events. This approach is thought to be especially appropriate when collection of data is expensive or time consuming (Oommen et al., 2011).

The impacts of these various sampling strategies on the validity of the outcomes of logistic regression have been widely explored (e.g., see Kamei et al., 2007; Oommen et al., 2011). In this regard, both Kamei et al. (2007), and Oommen et al. (2011), have suggested that the over-sampling (creating dummy examples of rare events), and the under-sampling (removing actual instances of common outcomes), procedures improve the predictive capability of logistic regression models; Oommen et al. (2011) reporting that over-sampling produced marginally more accurate estimates of the actual probabilities of events in the population.

Of course, there are always caveats to be introduced when dealing with such statistical procedures. Although a balanced outcome design in logistic regression has been noted to produce more powerful statistical outcomes (see Howell & Davis, 2011; Weiss, 2004), Oommen et al. (2011) noted that the predicted probability of an outcome derived from a sample undergoing logistic regression is closest to the true population probability when the sample has the same outcome distribution as the population: “Therefore, in probabilistic modelling using MLLR, it is important to develop a sample that has the same class distribution as the original population rather than ensuring that the classes are equally sampled.” (Oommen et al., 2011, p. 118). Thus, the aim of sampling should be to replicate the population outcome class ratio in the sample. In the case of under-sampling, the actual number of more common events from the sample selected for inclusion in the analysis will depend on the population ratio of the rare and common events, and on the number of less common outcomes sampled. For example, if the known population class distribution is 100:1, and the research has sampled 20 instances of the rare outcome, it should include 2000 randomly selected cases of the more common outcome. While there will be a problem from having a case bias, this will be reduced as the sample size increases (Oommen et al., 2011), indicating, again, the need for large samples.

Given these considerations, two problems immediately present themselves for most research studies. Firstly, the need for samples in the thousands may be difficult to achieve for most researchers, although, with the development of electronic databases (see Dworzynski et al., 2007; Ford et al., 2012), these participant numbers are becoming easier to accommodate. However, in much other research this issue of sampling bias and class bias remains a problem. Secondly, in many instances there are not clear cut population prevalence figures on which to base a sampling strategy, although in stuttering research these figures are largely agreed upon (see Craig et al., 2002; Howell, 2010; Mansson, 2000).

An example of how such a class imbalance problem can be tackled, without the collection of either unfeasibly large data sets, or the adoption of potentially questionable sampling strategies, is provided by Howell and Davis (2011). In this study, the factors that predicted the persistence of stuttering in later life were examined. Howell and Davis (2011) conducted research on subjects from the age eight until their teenage years in order to predict the persistence or recovery from stuttering using logistic regression. The use of an older population in this study not only filled a research gap, but also avoided a potential class imbalance problem. As regards the numbers of younger children who stuttered before the age of eight, the percentage of these children who recover by that age is four times larger than the percentage of in whom the stutter persists (Yairi & Ambrose, 2004). Thus, any classification procedures, such as logistic regression, that uses any sample based on this population of younger children, would have higher numbers of recovered stutters than persistent stutters. Consequently, any model developed to predict stuttering may have a poor fit to the data due to the power problems arising from class imbalance. However, the recovery- and persistence-rate of stuttering are more or less balanced in children at older ages, and the use of this group by Howell and Davis (2011) produced a more case balanced model for predicting risk for onset of stuttering, with consequent increase in the power of the analysis. Thus, solutions of class and sampling imbalances can be sought in the research design, as well as in sampling correction procedures.

2.3.3. Selecting outcomes

As noted above, logistic regression is used to understand the predictors of a dichotomous outcome event. In many cases, that outcome event is easily categorized into classes of having occurred, or not having occurred; for example, the occurrence of a heart attack, or of going bankrupt, are relatively easily discerned and coded as either having happened, or not having happened. Once this categorisation has been achieved, the predictors of that outcome can be studied (see Fleck et al., 2005; Greenhouse et al., 1995; Khan et al., in press, for examples of these types of variables).
In other cases, the outcome may be treated as dichotomous, but, in fact, it derives from the ‘censoring’ of continuous data; that is, a cut-off criteria has been produced, and the data recoded from continuous to categorical at the cut-off point. In these cases, the situation in choosing the outcome variable may be more complicated (see Greenhouse et al., 1995). Of course, some continuous outcomes translate relatively easily into a dichotomous event. These cases are most often concerned with measures for which well established cut-off points for the presence of an event have been developed. The presence or absence of high blood pressure is one such example, where a systolic pressure of greater than 140 mm/Hg is considered to be high (e.g., Chan, 2004). Similarly, where there are well documented points at which an event can be said to occur, based on an individual’s psychometrically-measured characteristics, then the coding of an event is relatively straightforward. The presence or absence of depression according to the Hospital Anxiety and Depression Scales (Zigmond & Smith, 1983) is one such example, although there will always be room for some debate about the exact cut-off points that should be employed (e.g., Bjelland, Dahl, Haug, & Neckelman, 2002).

The situation concerning the identification of outcomes in terms of stuttering research is, likewise, not necessarily clear cut, and, in this aspect, the area is no different from many other areas of psychology (see Greenhouse et al., 1995). The presence or absence of stuttering is, perhaps, the variable that is of most interest in the context of logistic regression in this field, both in terms of attempting to identify the factors that relate to its onset (e.g., Reilly et al., 2009), and that relate to its persistence over time (e.g., Howell & Davis, 2011) or in the face of intervention (Jones et al., 2000). The difficulty is how to define the presence of stuttering, and even a cursory examination of the literature reveals a range of ways of defining the presence of stuttering. Dworzynski et al. (2007) used parents’ ratings of stuttering, and Reilly et al. (2009) similarly used agreement across raters, including speech therapists and parents, whereas, Howell and Davis (2011) employed a measure relying on agreement between standardized measures.

Clearly, different means of identifying the presence or absence of the event in question will make difficult any comparison across different studies employing logistic regression to analyse putatively the same outcome. Moreover, the selection of different means by which to classify an outcome could produce different results from the logistic regression. The impact of particular decisions as to which cut-off points to employ in a psychometric measurement (see Howell & Davis, 2011), or the degree of agreement to accept as indicating the presence of stuttering (see Reilly et al., 2009), may have important implications for the results of the analysis, and a failure to attend to this aspect of the procedure may limit the usefulness or generality of the results (see Karp, 2004; Weiss, 2004).

2.3.4. Selecting potential predictors

Another aspect to consider in the development of a logistic regression study concerns the selection of which variables to analyse as potential predictors of the outcome (see Yairi and Ambrose, this issue, for such proposals). Ultimately, this can only be achieved by a careful study of the literature in relation to the outcome, in order to ensure that the full range of potential predictors is included. However, there are a number of pitfalls in selecting predictor variables that can lead to the presented logistic model appearing to explain greater (or lesser) amounts of variance than it actually may explain in reality.

Clearly, the results of any logistic regression will depend on the variables selected as potential predictors – put crudely, if a variable is not selected for analysis, then it cannot feature in the final model. This may make comparison across studies employing this technique more difficult; a brief scan of various logistic regression-based studies concerning the factors leading to stuttering onset or persistence reveals that different studies do employ different variables. For example, Reilly et al. (2009) employed three child variables, one parental health variable, ten family-demographic variables, and one socio-economic status variable in their analysis of stuttering onset. In contrast, Howell and Davis (2011) only employed seven potential predictor variables in their analysis of stuttering persistence. In the light of this, the fact that different factors emerge across the studies is not surprising, and the absence of several factors from the analysis is noted by both sets of authors.

However, the choice regarding whether or not to include factors in the initial data set can impact on the results, especially when the final equation is to be used in order to make quantitative predictions about the likelihood of a change in one factor impacting the outcome. As with many forms of regression analysis, logistic analysis attempts to demonstrate the impact of a predictor on the outcome when the impacts of the other variables on that outcome are controlled (see above). Obviously, only those variables present in the model can be controlled for – and the choice to include six, or seven, or eleven, variables will impact on this result. Moreover, if interaction terms between the variables are to be considered, then the omission of some variables could potentially have major impacts of the results.

Unfortunately, the solution is not simply to include as many variables as possible, as the inclusion of variables that are unrelated to the outcome in question, has the tendency to inflate the apparent predictive validity of the final model – an effect known as ‘Freedman's Paradox’ (Freedman, 1983; see Lukacs et al., 2010, for a discussion). Thus, the initial choice of predictors should be given equal consideration as the choice of outcomes.

There is no fool-proof way to tell that the set of predictors that have been chosen are appropriate, but a number of rules-of-thumb can show that the choice is reasonable. For example, if the specificity (i.e., the degree to which the predictors correctly identify individuals not showing the particular outcome – true negatives), and sensitivity (i.e., the degree to which the predictors correctly identify individuals showing the outcome – true positives), of the model are both above 80%, then it is likely that the chosen predictors have validity. Also, from the stepwise procedure mentioned earlier, the set of predictors (at least from the range initially studied) that are useful can be identified. Of course, this does not guarantee that a comprehensive set of the correct ones was there to start with.
There may well be constraints acting on any particular study that lead to bias in the selection of the data used for the analysis. One potential constraint is the sample size, which limits the number of variables that can be studied. There is some debate as to the number of participants per variable that are needed, however, Agresti (2007) suggests that 10 participants are needed for every variable studied; a suggestion that is based on some statistical evidence confirming the reliability of logistic regressions performed on different numbers of events per variable (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). This obviously places some constraints on the number of variables that can be employed in a study, although it should be noted that most studies of stuttering outcome using logistic regression do follow this rule.

Another source of selection bias in the variables that are studied is that of missing data, where the presence of missing data in the sample can drive down the sample size if those participants with missing data are excluded, or can lead to the exclusion of certain variables from the analysis if large amounts of data are missing. Unfortunately, both of these outcomes can lead to bias in the variables selected that may be highly important as it will leave the sample as self-selected – that is, comprising only those individuals who chose to supply certain data, or only that data which is readily supplied by the sample, as well as other reasons why the other data are missing. Dublin and Rivers (1990) discussed this issue in relation to logistic regression, and highlighted several procedures that can be employed to counter this problem.

Finally, in addition to selection bias effects from these sources, the selection of variables is also constrained by the properties of the data that are collected. For example, predictor variables that are related to one another (i.e., that show collinearity or multi-collinearity), or predictor variables that have excessively influential observations (outliers), will impact adversely on the results of a logistic regression. Particularly, in small or moderate samples, collinearity can result in overall levels of significance from the logistic regression when individual predictors are not in themselves predictive of the outcome, or in the degree of relationship between a predictor and the outcome being incorrectly established.

3. Alternatives to logistic regression

It should be noted that there are a number of alternatives to the use of logistic regression. Some of these procedures are quite familiar to most researchers (e.g., group testing techniques, such as t-tests, standard simple- and multiple-regression procedures, and discriminant analysis), and some are perhaps less familiar (such as the use of decision trees, and neural net algorithms). There are advantages and disadvantages to employing these analytic procedures in the situations described here. The former procedures (e.g., t-tests, multiple-regression) are rather simpler to apply and understand, but have limitations in both their applicability and in their informativeness. The latter set of procedures (e.g., neural networks), can often produce informative and appropriate results, but are difficult to employ and often require large data sets. However, some brief mention of these alternatives may help to place the advantages of logistic regression techniques into relief.

3.1. Group testing

A number of studies have divided the sample into those who stutter and those who are fluent, and then compared these two groups in terms of factors such as utterance length and syntactic complexity (e.g., Yaruss, 1999), and emotional reactivity (e.g., Karrass et al., 2006). Differences between the groups in terms of these variables are then used to identify factors that may distinguish between those who stutter and those who are fluent. Although this is a readily understandable technique, and one that is easy to interpret, it suffers from a number of problems. Firstly, it reverses the relationship between the outcome and the predictor variables, making the outcome of interest into the independent variable, and the predictors into the dependent variable. Of course, there are also interpretative problems in such quasi-experimental designs; they may give the illusion of a causal relationship, but they are no more than a correlational procedure – there being no random allocation into groups. Another problem with this approach is that, while it can show variables on which two outcome-defined groups differ from one another, it is not always straightforward to derive the strengths of the relationships between the variables and the outcome in question.

3.2. Regression procedures

Obviously, a technique that comes to mind in this context is a simple- or a multiple-regression, in which the overall strength of the impact of a predictor can be seen on an outcome, and that relationship can be characterised by an equation of the form: outcome = b0 + b1 predictor, where b0 and b1 are constants. This form of analysis has been employed in stuttering research in order to explore predictors of stuttering relapse (e.g., Hancock and Craig, 1998), and of perceptions of stuttering (e.g., Ginsberg, 2000). Although the two examples, noted above, used outcomes that were continuous in nature (percentage of syllables stuttered, and perception of stuttering scales), it is quite possible to use a dichotomous variable as an outcome (see Howell, 1997). However, the use of continuous variables may not be optimal in clinical contexts where: (1) clear decisions about outcomes are sought, (2) such continuous scales may be too time consuming to administer, and (3) the scales may be of limited reliability compared to clear clinical outcomes, such as the presence or absence of an event.

However, more important limitations exist on the use of these simple regression techniques if dichotomous outcomes are used. In particular, as noted above, these situations often tend to produce violations of the heteroscedasticity assumption required for regression (i.e., the variance in the outcome is the same at all measured levels of the predictor; see Howell, 1997; although see Cox & Wermuth, 1992, for discussion of when this might not be a problem). Additionally, regression
procedures will produce predictions that involve unreal numbers for these contexts; that is, the prediction must always be either the presence or absence of an outcome (i.e., a zero or a one). Simple- and multiple-regressions will produce predictions of outcomes between these possible outcomes, and of outcomes that are greater than one.

3.3. Discriminant analysis

A discriminant analysis attempts to predict an individual’s membership of one from a number of mutually exclusive groups based on their scores from predictor variables. For example, whether it is possible to predict that somebody will stutter based on whether they have attended an intervention programme, or on any number of demographic and personal characteristics. Discriminant analysis is most often used when there are only two categories of outcome; but it can also be used for multi-factor outcome categories. In this latter capability, it has an advantage over logistic regression, which is always applied to dichotomous outcomes. However, logistic regression requires fewer assumptions about the data, and is a more robust statistical test in practice.

3.4. Complex techniques

It should also be noted that there are several sophisticated alternatives to logistic regression techniques available, such as decision trees, and neural networks. Several of these techniques may have some advantages over logistic (Ayer et al., 2010; Khemphila & Boonjing, 2010; Tu, 1996, for discussions of their relative merits in medical contexts). Although there may well be advantages to the use of these procedures, there are also some pertinent drawbacks. In many cases, the criteria for judging whether the relationships identified between variables are statistically significant are not clear. These procedures are also often best suited to the analysis of very large datasets, which are often not collected in stuttering research. Finally, these procedures have not, as yet, been applied to any great level in stuttering research but they should be.

4. Summary

Logistic regression offers an increasingly employed and recognized approach to allow prediction of clinical relevant dichotomous outcomes. It presents several advantages over more traditional approaches to the analysis of such data (e.g., t-test and multiple-regression), and is better explored in this context than newer data analytic procedures (e.g., neural nets). However, there are a number of important considerations that need to be taken into account when planning research that will be analysed by logistic means. These considerations include: the use of a large sample size; the use of limited numbers of, and well-thought through, predictors; and an assessment of the likely ratio of positive and negative outcomes. However, once these issues are overcome, the outcomes of logistic regression can be very illuminating and practically useful in establishing the relationships between events in stuttering research.

CONTINUING EDUCATION

Logistic regression for risk factor modelling in stuttering research

QUESTIONS

1. The steps in a logistic regression are:

(a) Obtain the dependent and independent variables and plot them against one another.
(b) Calculate the mean for the independent variable and compare them across treatment and no-treatment groups.
(c) Calculate the probabilities of outcomes at each value of the predictor; convert this probability to an odds ratio; and take the natural logarithm of the odds ratio.
(d) Equivalent to the steps in an Analysis of variance.

2. The preferred technique for establishing factors that increase the risk of stuttering is:

(a) T test.
(b) Linear regression.
(c) Binary logistic regression.
(d) Analysis of variance.
(e) Chi square.

3. Alternative techniques to logistic regression:
(a) There are no alternatives in situations where logistic regression is used.
(b) There are alternatives in situations where logistic regression is used but these are always more complicated.
(c) There are alternatives in situations where logistic regression is used but these are always simpler and have shortcomings.
(d) There are alternatives which have no disadvantages.
(e) There are alternative which have disadvantages.

4. Logistic regression has been applied to:
   (a) Diagnosis of stuttering and types of stuttering.
   (b) Prognosis of stuttering and severity of stuttering.
   (c) To follow treatment outcomes and diagnosis of stuttering.
   (d) Types of stuttering and severity of stuttering.
   (e) Severity of stuttering and to follow treatment outcomes.

5. The optimal conditions under which to perform a logistic regression are:
   (a) With an equal number of cases for each type of outcome.
   (b) With large numbers of predictor variables.
   (c) With unequal numbers of cases for the different types of outcomes.
   (d) With many types of outcome.
   (e) When every variable is normally distributed.

References


**Phil Reed** is Chair in the Department of Psychology Swansea University. Phil’s research interests include: Behaviour Analysis, Learning and Memory, Schedules of Reinforcement, Behavioural Interventions. Phil has held appointments for the Department for Education (Special Educational Needs), the Department of Health, and the Children in Wales Policy Council.

**Yaqiong Wu’s** research interests are stuttering, sensori-motor control and motor learning in stuttering. She graduated on the M Res course on speech and language at University College London in 2011 Her mentor and project supervisor was Professor Howell. She has returned to China where she continues her research interests.