Back to Session Subjects

A BETTER WAY TO ANALYSE AND INTERPRET HELICOPTER HANDLING QUALITIES AND WORKLOAD RATINGS.[1]

R Bradley and W M Maclaren

School of Computing and Mathematical Sciences,
Glasgow Caledonian University,
G4 0BA
UK
r.bradley@gcal.ac.uk
w.maclaren@gcal.ac.uk

Keywords: Handling qualities, workload, OLR

Abstract: The paper describes the application of new analysis methods to handling qualities ratings data returned by pilots during flight trials and piloted simulation. The method offers the potential to quantify the effect of variations in, for example, pilot training and experience, atmospheric conditions or control law configuration. Flight trials and piloted simulation are expensive and time consuming to mount and are conducted with diligence and expertise by experienced personnel. Particular care is taken to return ratings that are consistent and compliant with the set criteria. It is important, therefore, that this professional diligence is continued through to the analysis and interpretation of ratings returned by the pilots. The statistical method used in the analysis is *ordinal logistic regression*, which allows the specifying and fitting of regression relationships between ordered categorical response variables and explanatory variables. The response variable in this context is the handling qualities rating (HQR) on an ordered categorical scale of 1 to10 (including fractional levels if present in the data) and the explanatory variables are the selected experimental factors - such as wind speed, manoeuvre aggression or pilot identity - believed to influence the rating.

1. INTRODUCTION

The Cooper-Harper pilot rating scale [1] is a widely accepted measure of helicopter handling qualities and is extensively employed in flight tests and piloted simulations. Through a structured debriefing based on a decision tree involving the aircraft's characteristics, task performance and workload the pilot returns a rating on a scale of 1 - 10. The skill of the test or flight simulation engineer is to guide the pilot through the rating dialogue in a manner which is designed to give repeatable and consistent results. As a further guard against pilot variability, the opinion of several pilots may be recorded and the range of ratings presented as an average and max/min range [2]. Despite these procedural precautions it is often the case that pilots return half ratings or ratings qualified with '+' or '–'. In any case, arithmetical manipulation of handling qualities ratings (HQRs) is not appropriate since this process introduces the likelihood

[1] *Prepared for the 32nd European Rotorcraft Forum, Maastricht, The Netherlands, September, 2006.*

of a non-integer value for which the associated definition is imprecise. Such ratings are properly regarded as ordered categorical data and recent developments [3] have introduced modern statistical analysis techniques for the analysis of such data. The purpose of this paper is to illustrate the application of one such technique, Ordinal Logistic Regression (OLR), to HQR data from a number of piloted flight simulation trials. A brief description of OLR, is given in the appendix to this paper and several illustrations of its application to HQR data are given in the sections which follow this introduction.

The Cooper Harper scale is typically employed in the comparative evaluation of different aircraft - or aircraft configurations - but when a single aircraft is engaged to investigate the effect of external influences on the returned HQRs, the ratings may be interpreted as a measure of workload. A study of this type arose when, in an update to CAP 437: Offshore Helicopter Landing Areas - Guidance on Standards published by the Civil Aviation Authority (CAA) [4], a programme of work was commissioned by CAA at BMT Fluid Mechanics, supported by QinetiQ Ltd. and Glasgow Caledonian University to define a maximum permitted level of turbulence around offshore platforms. A full description of the whole programme is provided by Rowe [5, 6]. One strand of this research involved a programme of piloted simulation trials carried out on the QinetiQ Advanced Flight Simulator (AFS) using a generic helicopter model configured to resemble a Sikorsky S76 - a type common in North Sea operations. The flight task involved approaches to, and establishing a stable hover 10 ft above, a representation of the helideck on the Brae-A platform in various wind speeds and directions. The dedicated hover task involved wind speeds of 15, 25, 35, 50 and 60 kt, and wind directions of 182°, 271°, 320° and 358° - coinciding with obstructions which were potential sources of increased turbulence. They were respectively: unobstructed (as a baseline), derricks, cranes and exhaust stacks. The 3D turbulence field required for the simulation trial was derived from data from the BMT Boundary Layer Wind Tunnel using a 1:100 scale model of the Brae-A platform. This particular study, therefore involved: (i) constructing a model of the Brae-A platform, (ii) carrying out wind tunnel experiments for a range of orientations, (ii) capturing and collating the data of the 3 directions of turbulence, (iii) integration of the turbulence measurements into the simulation environment, (iv) configuring a generic helicopter model to resemble a S76, (v) mounting the programme of piloted trials (3 pilots were used) and (vi) debriefing after each run to obtain a returned HQR. This brief summary of activities emphasizes the range of expertise and scale of resources that must be deployed to obtain, in this case, approximately 60 HQRs. This situation will not be unfamiliar to simulation engineers: piloted simulation and flight tests are complicated and expensive to implement. There is every reason, therefore, to ensure that the maximum amount of information is extracted from what is usually a relatively small number of resulting HQRs.

The HQR data from this research have recently been analysed in substantial detail by Bradley and Maclaren [7] using OLR techniques. A surprising result from the analysis was that the variation in returned HQRs between pilots was tested to be more significant than the variation in wind direction. Also there was no significant difference between any of the wind directions except when it impinged on the derricks. Figure 1 shows the probability of returned HQRs for the three pilots as a function of wind speed for this direction. This figure is easily interpreted. For example, for a wind-speed of 15kt, Pilot 1 returns an HQR of 3 with a probability of

0.5 and an HQR of 4 with a probability also of 0.5; whereas for a wind-speed of 25kt the HQRs of 4 and 5 are returned with probabilities 0.8 and 0.2 respectively. These two features: (i) the testing for the significance of the explanatory variables and (ii) the estimation of the probability distribution of returned ratings, are believed to be an important advance in the analysis of HQRs.

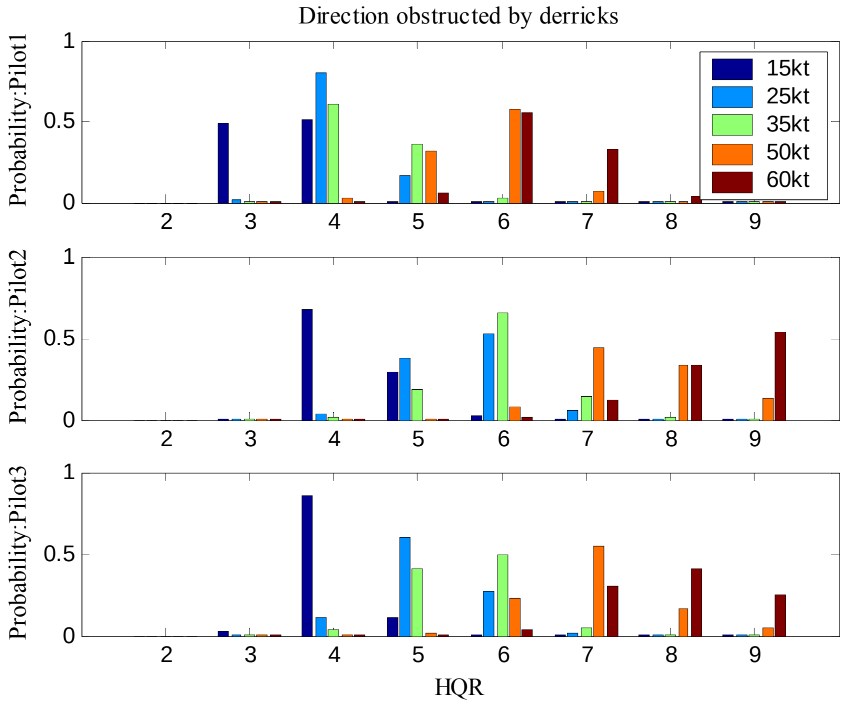

Figure 1. Probabilities of returned HQRs for wind direction impinging on derricks.

Following the success of the ORL analysis of the BRAE-A helideck data, HQR data from piloted simulation trials were revisited with the aim of testing the efficacy of the approach on small data sets. The remaining sections of this paper consider the HQRs collected during three piloted trials, designated TWIN1, CONDVAL and TWIN3, on the Advanced Flight Simulator (AFS), at DERA (now QinetiQ), Bedford. During the TWIN1 [8] trial, in an investigation of the effect of different roll attitude bandwidths on pilot workload and task performance, the roll damping and roll control sensitivity were varied between three Conceptual Simulation Model (CSM) [2] configurations. The CSM is a simplified model of the dynamics of a helicopter and was specifically developed to investigate the relationship between response characteristics and handing qualities criteria. The three resulting configurations designated T1C1, T1C2 and T1C3, in descending order of bandwidth, are intended to respectively represent; an Active Control Technology (ACT) Lynx with perfect decoupling, a datum configuration corresponding to the in-service Lynx, and a degraded case. The CONDVAL trial [9] presented the first piloted assessment of the AFS configured as a Lynx using the high-fidelity helicopter simulation model HiFiSim Lynx. Its improved main rotor modelling provides increased fidelity for simulating flying qualities using the AFS. The objective of the trial was to provide pilot stick activity data, for workload and handling qualities studies. These couplings were not present in the earlier TWIN1 data as the CSM does not include a representation of the main rotor dynamics. The aim of the TWIN3 [10] trial was to investigate the effect on Lynx handling qualities of varying the authority of the flight control series actuator when performing aggressive manoeuvres. The baseline configuration, designated T3C1, corresponds to the standard HELISTAB [2] Lynx and its Automatic Flight Control System (AFCS) with approximately 13% authority. The second case considered, T3C2, was an un-augmented Lynx with the AFCS switched off. Finally, the third configuration, T3C3, was a Lynx with its AFCS at full (100%) authority.

The HQR data for the TWIN1, CONDVAL and TWIN3 trials have been taken from reference 11 and were originally supplied by DERA for a study on metrics for workload prediction.

2. THE TWIN1 TRIAL

The first example to be considered is data from the piloted simulation of a modified ADS-33 Slalom, consisting of two Slalom elements separated by a straight section of track, flown by two experienced test pilots. The returned HQRs are shown in Table 1. Recall that the three configurations T1C1, T1C2 and T1C3, are in descending order of bandwidth.

Run	**Configuration**	**Pilot**	**HQR**
1	T1C1	P1	3
2	T1C1	P1	4
3	T1C3	P1	4
4	T1C3	P1	5
5	T1C3	P1	6
6	T1C2	P1	5
7	T1C2	P2	5
8	T1C1	P2	4
9	T1C1	P2	5
10	T1C1	P2	6
11	T1C3	P2	7
12	T1C2	P2	4
13	T1C3	P2	5
14	T1C3	P2	6

Table 1. AFS TWIN1 Slalom Trial

The HQRs vary in this case from 3 to 7 and are integer values. The regression model contains two factors: configuration and pilot. The baseline includes the effect of configuration T1C1 and pilot P1 (see appendix for an explanation of the terminology). Then there are three indicator variables to capture the variation from the baseline: two for the configurations T1C2 and T1C3, and another one for the pilot P2. The estimated probabilities of awarded HQRs from the OLR are shown in figure 2.

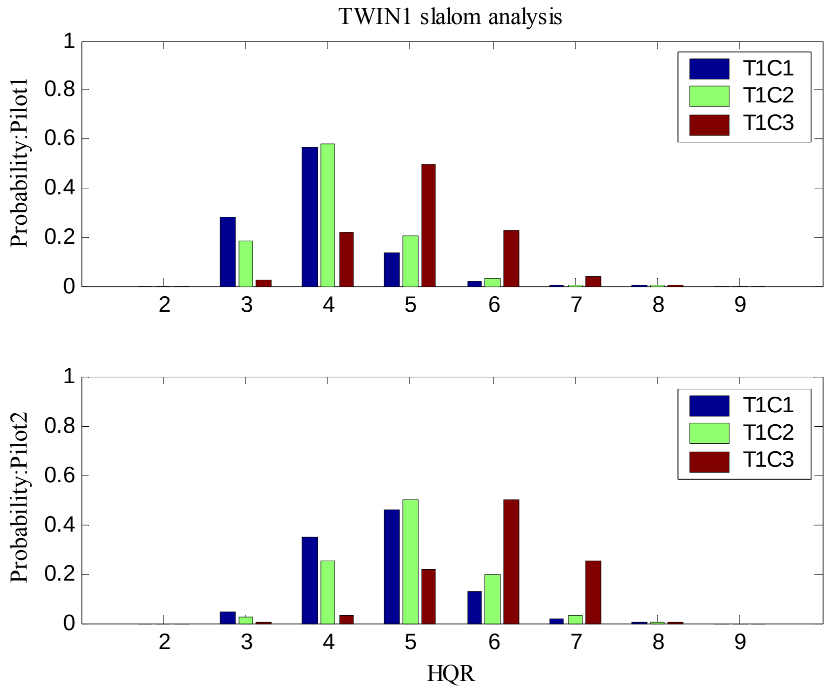

Figure 2. Estimated Probabilities in AFS TWIN1 Trial Slalom Runs.

For both pilots, configuration T1C3 attracts the highest rating, but there does not appear to be a clear distinction between T1C1 and T1C2. For this configuration, T1C3, pilot P1 the most probable HQR is 5, with probability approx. 0.5, but HQR 4 and 6 may be awarded, each with probability 0.2. The shift to the right of the distribution for pilot P2 compared to that for P1 suggests that pilot P2 systematically returns a rating higher than P1. However, tests for the inclusion of the factor 'Pilot' alone shows it not to be significant at the 5% level (P=0.215) which may be interpreted as indicating that there is no difference between the two pilots. The situation for the effect of 'Configuration' is not clear cut since the tests for its inclusion alone gives P=0.612 and P=0.077 for T1C2 and T1C3 respectively. The latter case is close to the 5% level and, indeed, when both factors: Pilot and Configuration, are included together the value for T1C3 becomes 0.042. This situation reflects a borderline case and attracts the terminology that the data are suggestive of a connection between variation in configuration and HQR probabilities. The data, therefore, may be interpreted as supporting the view that there is no difference between the pilots and suggesting that higher HQRs arise from TW1C3 compared to TW1C1 and TW1C2.

The data in table 1 consists of 14 samples involving 2 pilots, 3 configurations and 5 rating categories. Such data would be regarded by the simulation fraternity as a successful trial but in statistical terms the sample is small and caution is required in interpreting the results of the significance tests. Nevertheless, the derived probability distributions are valid and provide a revealing insight into the award of ratings.

3. THE CONDVAL TRIAL

In this trial, three test pilots performed four manoeuvres of which we consider here the Bob-up/down and Accel/decel MTEs at three specified levels of task aggression: low, moderate and high. The data from the Bob-up/down trial are shown in Table 2.

Run	Aggression	Pilot	HQR
1	L	M	4
2	L	M	4
3	H	M	6
4	H	M	6
5	M	O	4
6	L	O	4
7	L	O	4
8	L	O	4
9	L	T	4
10	L	T	4
11	M	T	5 (-)
12	M	T	5 (-)

Table 2. AFS CONDVAL Bob-up/down Trial

The HQR data ranges from 4 to 6 in 3 categories. The category 5- is regarded simply as a category between 4 and 6. Again, in the analysis there are two factors: this time 'Pilot' and 'Aggression' with baseline levels M and L respectively. The regression contains four indicator variables corresponding to the two additional levels for each of Pilot and Aggression'. The estimated probabilities from the OLR of awarded HQRs in this trial are shown in figure 3. An inspection of the probability distributions in figure 3 suggests that there is no great difference between Pilots M and O but Pilot T appears return a higher rating. They all rate the high aggression (H) task as involving greater workload compared to the low (L) and medium (M) aggression tasks - between which there appears little difference. However, in table 2 only Pilot M performs the high aggression task so that the predicted probabilities for this level of aggression for Pilots O and T are obtained by extrapolation using the regression model. For example, the probability of 1.0 that Pilot T will return HQR of 7 for the high aggression case is initially surprising since the table contains no entries for this value. These interim conclusions have not yet been tested for significance but there is a greater need for caution in this case as there are only 12 samples involving 3 pilots, 3 aggression levels and 3 rating categories. Further, as has been noted, there is no data for some combinations of the Pilot / Aggression factors. It should be noted that the method does not impose any *a priori* ordering on the levels L, M and H. Any relationship between the task aggression and the HQR probabilities is derived from the data.

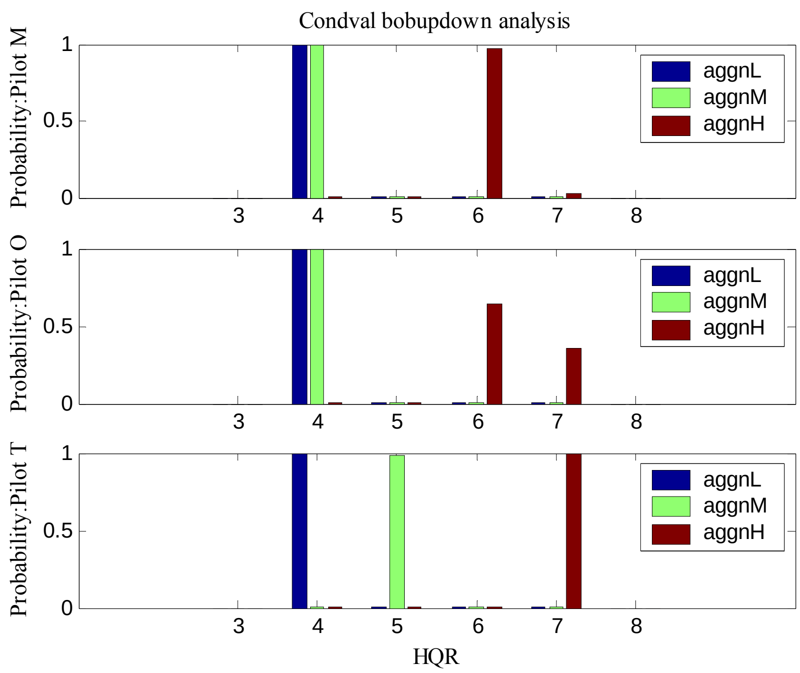

Figure 3. Estimated Probabilities in AFS CONDVAL Bob-up/down Trial

The second case is the Accel/Decel manoeuvre - the data from which are shown in table 3.

Run	**Aggression**	**Pilot**	**HQR**
1	L	M	4
2	M	M	4
3	M	M	4
4	H	M	6
5	H	M	6
6	L	T	3
7	L	T	3
8	M	T	4
9	M	T	4
10	H	T	5+
11	H	T	5+
12	L	O	4
13	L	O	4
14	MH	O	4
15	M	O	4
16	H	O	5
17	H	O	5

Table 3. AFS CONDVAL Accel/Decel Trial

The HQR data ranges from 3 to 6 in 4 categories. The category 5+ is regarded simply as a category 5 for this exercise - a later example will illustrate how it may be included as a separate category between 5 and 6 if required - and if there are sufficient data to support it. Again the two factors are Pilot and Aggression with baseline levels M and L respectively. This time, however, there are five indicator variables: two for the additional two levels of the Pilot factor and three for the Aggression factor due to the single occurrence of the MH (medium-high) level. The estimated probabilities from the OLR of returned HQRs in this trial are shown in Figure 4.

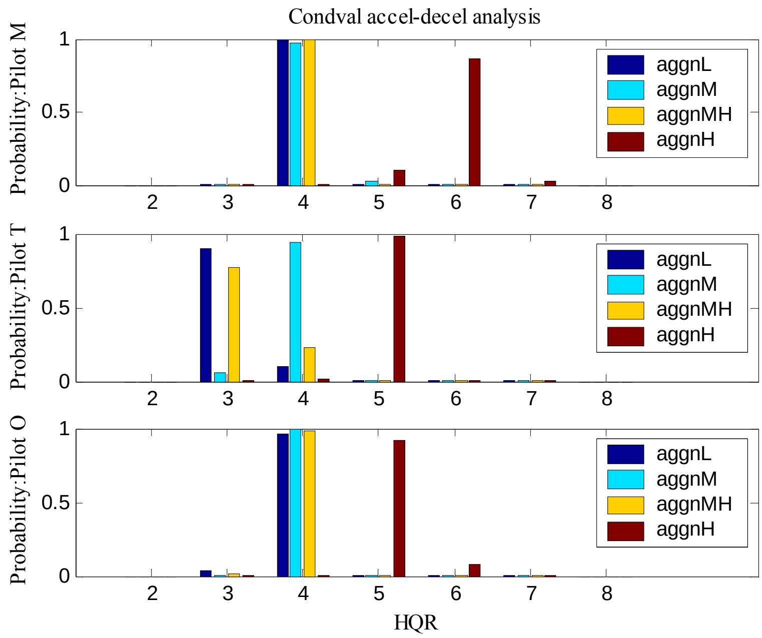

Figure 4. Estimated Probabilities in AFS CONDVAL Accel/decel Trial

An inspection of the probabilities in fig. 4 again shows little difference between pilots and an increase of workload for the high aggression manoeuvre compared to the low and medium. Note that the single MH rating of 4 by pilot O is reflected in the distribution with a probability of 1.0. The regression model uses all of the data to estimate probabilities for pilots M and T based on how they rated other runs. Further, Pilot O returns HQR =4 for levels L, M and MH of aggression and therefore the only information available on the level MH does not distinguish it from level M and L. As was observed above, there is no *a priori* relationship between ratings and aggression levels. The initially surprising consequence of this situation is that in figure 4, Pilot T is predicted to return ratings for MH which are more frequently lower than those for M.

These interim conclusions regarding no significant pilot variability and increase of HQR with level of aggression remain to be tested for significance. The deletion of the single sample for aggression level MH is probably advisable; in fact the predicted probabilities for the edited data set, shown in figure 4a, vary little from those in figure 4.

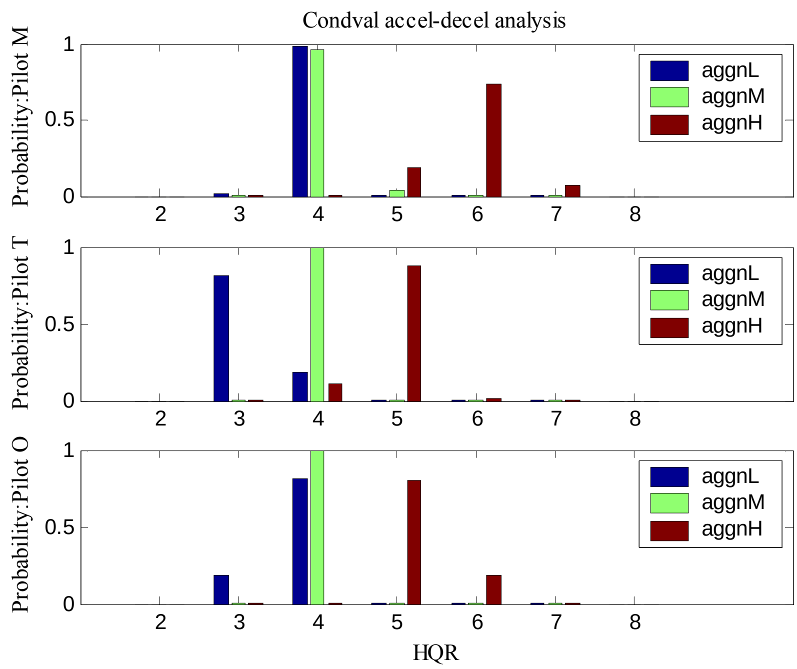

Figure 4a. Estimated Probabilities in AFS CONDVAL Accel/decel Trial (case MH deleted)

Comparing the analysis of the Bob-up/down and Accel/decel manoeuvres, it is clear that the classification of aggressiveness has been carefully judged since the manoeuvres in these vertical and longitudinal directions respectively give similar handling qualities responses.

4. THE TWIN3 TRIAL

In the TWIN3 trial, three vehicle configurations with Levels 1, 2 and 3 handling qualities characteristics (as determined by the ADS-33 criteria) were evaluated by three test pilots flying four manoeuvres. Again, here we consider two cases the Bob-up/down and Accel/decel MTEs. Recall that the configurations are Standard (T3C1), Unaugmented (T3C2) and Full authority (T3C3). The data from the Bob-up/down trial is shown in table 4. The HQR data range from 3 to 7 in 5 categories. The category 4- is regarded simply as a category 4 for this exercise. There are two factors in the OLR analysis: Pilot and Configuration with baseline levels P1 and T3C1 respectively. There are indicator variables for T3C2, T3C3, P2 and P1 in the analysis. The estimated probabilities resulting from the OLR analysis are shown in figure 5.

Run	Configuration	Pilot	HQR
1	T3C1	P1	4
2	T3C1	P1	4
3	T3C2	P1	5
4	T3C2	P1	5
5	T3C3	P1	3
6	T3C3	P1	3
7	T3C3	P2	4
8	T3C3	P2	4
9	T3C2	P2	5
10	T3C2	P2	5
11	T3C1	P2	4(-)
12	T3C1	P2	4(-)
13	T3C1	P3	5
14	T3C1	P3	5
15	T3C1	P3	5
16	T3C2	P3	7
17	T3C2	P3	7
18	T3C3	P3	5
19	T3C3	P3	5

Table 4. AFS TWIN3 Trial Bob-up/down Runs

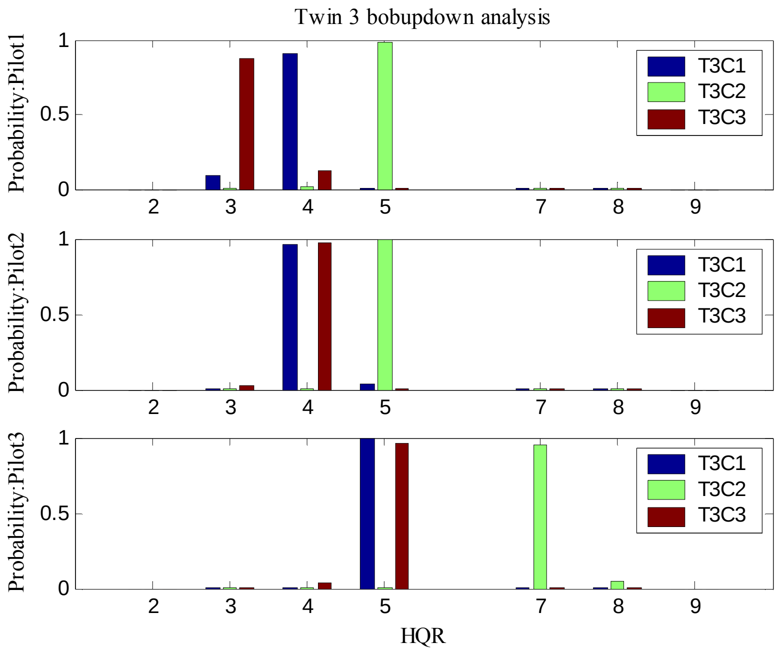

Figure 5. Estimated Probabilities in AFS TWIN3 Trial Bob-up/down Runs.

From an inspection of fig.5 there appears to be a clear distinction between the higher ratings of the T3C2 configuration (un-augmented) and the other configurations and pilot P3 would appear to return ratings at least one higher than pilots P1 and P2. These differences remain to be confirmed as significant by the OLR analysis as does the conclusion that there is no evidence of a significant difference in rating of the T3C1 and T3C3 configurations. The conclusion, by inspection at the present time, is that the experiment - that is the defined MTE - does not draw out a difference between 13% and 100% authority AFCS.

The data for the Accel/decel manoeuvre in the TWIN3 trial are shown in Table 5.

Run	**Configuration**	**Pilot**	**HQR**
1	T3C1	P1	4
2	T3C1	P1	4
3	T3C2	P1	5
4	T3C2	P1	5
5	T3C3	P1	2.5
6	T3C3	P1	2.5
7	T3C3	P2	5
8	T3C3	P2	5
9	T3C2	P2	5
10	T3C2	P2	5
11	T3C1	P2	4 (+)
12	T3C1	P2	4(+)
13	T3C1	P3	6
14	T3C1	P3	6
15	T3C2	P3	5
16	T3C2	P3	5
17	T3C3	P3	5
18	T3C2	P3	5

Table 5. AFS TWIN3 Trial Accel/decel Runs

The HQR data ranges from 2.5 to 6 in 4 categories. The category 4+ is regarded simply as a category 4 for this exercise but the rating 2.5 is retained as a category below 4 in this case. There is no rating of 3. The factors and indicator variables are the same as in the Bob-up/down case.

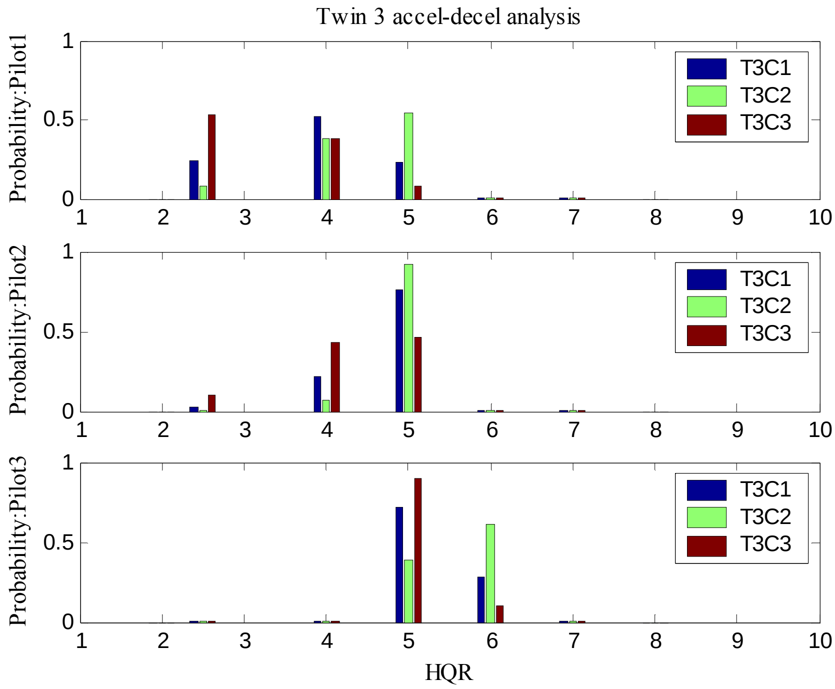

Figure 6. Estimated Probabilities in AFS TWIN3 Trial Accel/decel Runs

The estimated probabilities from the OLR analysis are shown in figure 6. Again pilot P3 appears to systematically award higher ratings than the other pilots. The un-augmented AFCS (T3C2) again appears to attract higher ratings but the situation is not so clear cut as in the Bob-up/down case. In particular, for pilot P2 there is not much to choose between any of the configurations. For the unusual rating of 2.5, the model predicts noticeable probabilities for pilot P1 and small values for P2 and P3. These latter two pilots do not, of course, award such ratings in the data but are given non-zero probability values via the fit to the regression model. The conclusions that (i) there is no significant difference in the ratings attracted by any of the configurations, but (ii) pilot P3 returns significantly higher ratings than his colleagues, remain to be confirmed.

As a final example we consider briefly the Side-step MTE from the TWIN3 trial. The data are shown in table 6. The HQR data ranges from 3.5 to 7 in 6 categories. The ordered categories are: 3.5, 4, 4.5, 5, 5.5, 7. There is no rating of 6. The unusual half-ratings cause no problems in the formulation of the OLR regression; they simply fall into the appropriate place in the ordering. What does cause a problem is that they introduce additional coefficients into the regression model for an already small data set and, further, the half-ratings are only employed by pilot P1 - who only uses half values. The factors and indicator variables, however, are the same as in the Bob-up/down case.

Run	Configuration	Pilot	HQR
1	T3C1	P1	4.5
2	T3C1	P1	4.5
3	T3C2	P1	5.5
4	T3C2	P1	5.5
5	T3C3	P1	3.5
6	T3C3	P1	3.5
7	T3C3	P2	5
8	T3C3	P2	5
9	T3C2	P2	5
10	T3C2	P2	5
11	T3C1	P2	4
12	T3C1	P2	4
13	T3C1	P3	4
14	T3C1	P3	4
15	T3C2	P3	7
16	T3C2	P3	7
17	T3C2	P3	5
18	T3C3	P3	5

Table 6. AFS TWIN3 Trial Side-step Runs

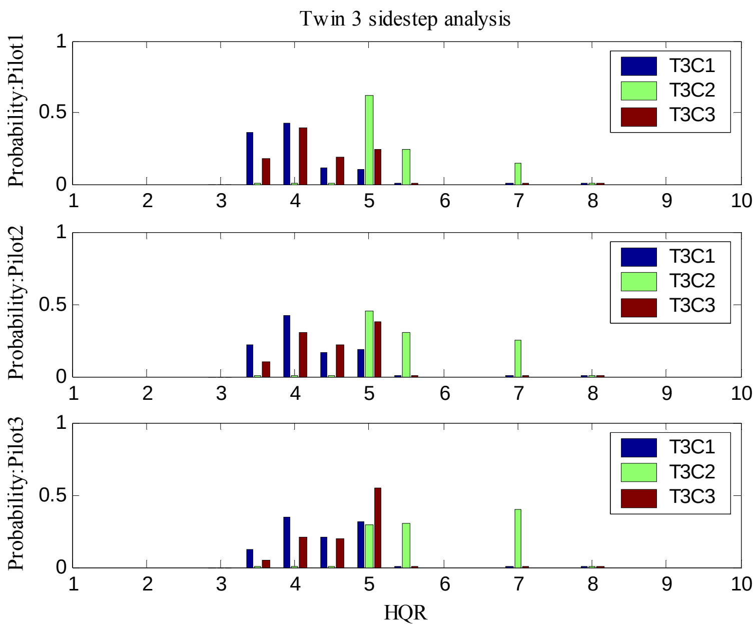

Figure 7. Estimated Probabilities in AFS TWIN3 Trial Side-step Runs

The estimated probabilities from the OLR analysis are shown in figure 7. The half-ratings can be seen to appear in the distributions for all the pilots as an artifact of the regression model. Pilot P3, again may be considered appears to award higher ratings than the other pilots and the un-augmented AFCS (T3C2) again appears to attract higher ratings. It should be clear that HQR data qualified by '+' or '−', or any other ordering notation are amenable to analysis by OLR - but all of these additional categories require additional data to support them and this consideration must be catered for in the experimental design.

5. CONCLUSIONS AND COMMENTS

Handling Qualities Ratings derived from the application of the Cooper-Harper scale are ordered categorical data and now that statistical methods are available in standard statistical packages, such as Minitab [12] and Genstat [13], they are to be preferred over arithmetical processing. For small data sets, however, the convergence criteria of different maximum likelihood algorithms can cause difficulties in achieving a unique solution to the regression model. Nevertheless, meaningful probability distributions for the HQRs may result.

Given adequate data, the ORL technique with HQRs as the response variable provides useful estimates of probability distributions for HQRs and significance tests for the explanatory variables.

Pilot variability in the award of ratings can be properly tested for significance - either confirming the consistency of results or flagging that it is an issue for the experiment.

The method supports and justifies the attention given to experimental design. For example, the guidelines on using more than pilot by Padfield [2] can be applied knowing that the resulting HQRs can be rigorously tested for significance.

The desirability of a data-set with test-points using as many of the total combinations of explanatory data as possible imparts additional validity to the analysis. That is, good experimental design is rewarded by statistical tests of increased validity.

The existence of methods than can extract meaningful and authoritative conclusions from HQRs should give confidence to those intending to establish expensive piloted simulation and flight tests.

6. ACKNOWLEDGEMENTS

The authors wish to acknowledge the Civil Aviation Authority, BMT Fluid Mechanics Ltd and QinetiQ Ltd. for their support in providing data for the analyses described in this paper.

7. REFERENCES

1. Cooper, G.E., Harper, R.P., The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities, Report No. NASA-TN-D-5153, April 1969.

2. Padfield, G.D., Helicopter Flight Dynamics: the theory and application of flying qualities and simulation modelling, 1995 (Blackwell Science, Cambridge, UK).

3. McCullagh, P., Regression Models for Ordinal Data, Journal of the Royal Statistical Society B. Vol 42, 1980 pp 109-142.

4. Anon., Offshore Helicopter Landing Areas - guidance on standards, CAP 437, Third edition published by the Civil Aviation Authority, London, October 1998.

5. Rowe, S.J., Howson D., Bradley R., The Response of Helicopters to Aerodynamic Disturbances around Offshore Helidecks, Royal Aeronautical Society Conference Helicopters in the Marine Environment, London, March 2001.

6. Rowe, S.J., Howson D., Turner G., A Turbulence Model for Safe Helicopter Operations to Offshore Platforms, 30th European Rotorcraft Forum, Marseilles, 2002, Paper 56, pp1-10.

7. Bradley, R., Maclaren, W.M., Ordinal Logistic Regression Analysis of Flight Task Ratings. Aeronautical Journal (to be published).

8. HOWELL, S.E., Preliminary Results from Flight and Simulation Trials to Investigate Pilot Control Workload in Slalom Manoeuvres, Flight Dynamics and Simulation Department, Defence Research Agency, Bedford, UK, DRA/AS/FDS/WP95181/1, June 1995

9. PADFIELD, G.D., CHARLTON, M.T. and MCCALLUM, A.T. The Fidelity of HiFiLynx on the DERA Advanced Flight Simulator using ADS-33 Handling Qualities Metrics: Report on the CONDVAL Trial - Flying Lynx in Good Visual Environment, Flight Management and Control Department, Defence Evaluation and Research Agency, Bedford, DRA/AS/FDS/TR96103/1, December 1996.

10. CHARLTON, M.T. and HOWELL S.E. Trial Specification for AFS Simulation Trial TWIN3, Draft Version, Flight Management and Control Department, Defence Evaluation and Research Agency, Bedford, UK, Ref:

DERA/AS/FDS/2TG5/22/01T/97/08, November 1997.

11. MacDonald, C.A., The Development of an Objective Methodology for the Prediction of Helicopter Pilot Workload, PhD Thesis, Department of Mathematics, Glasgow Caledonian University, UK, January 2001.

12. Minitab Statistical Software, Release 12, User's Guide 1998. Minitab Incorporated, PA.

13. Gilmour, A., et al. ASReml User Guide, Release 2, VSN International, ISBN 1-904375-21-9, 2006.

APPENDIX.

STATISTICAL MODELLING.

The ordinal logistic regression method of McCullagh allows the specifying and fitting of regression relationships between ordered categorical response variables and explanatory variables. The response variable in this context is the HQR on an ordered categorical scale of 1 to10 (including fractional levels if present in the data) and the explanatory variables are the selected experimental factors - such as wind speed, manoeuvre aggression or pilot identity - believed to influence the rating.

This technique builds a regression model of the probability γ_j of awarding a rating equal to, or below, each level j (j=1 to 9) and proceeds by successively introducing the next most significant explanatory variables. The general model takes the form

$$\log\left(\frac{\gamma_j}{1-\gamma_j}\right) = \alpha_j + \beta_1 E_1 + \ldots + \beta_k E_k \qquad \text{(A1)}$$

where $E_1 \ldots E_k$ are indicator variables, taking the value 0 or 1, for the k explanatory variables. This model predicts how the logarithm of the odds (logit function) of a rating j or less varies from a base value according to the presence of explanatory variables. The logarithm of the odds may not be a convenient quantity for interpretation of the results and even the probability γ_j of awarding a rating equal to, or below, a level j may not be familiar. We have found that a further step of differencing to calculate π_j the probability of returning a rating equal to j provides results of practical value. The probability π_j may be calculated from

$$\pi_j = \gamma_j - \gamma_{j-1}\ , j\text{=2…9}. \qquad \text{(A2)}$$

The coefficients, α_j and β_i, in the regression model are determined using the maximum likelihood criterion.

OLR can be done by a number of standard statistical packages such as the widely used Minitab software. This software primarily produces values for the regression coefficients α_j and β_i and associated tests of significance. From these coefficients, the γ_j may be subsequently calculated. The π_j are then easily found by differencing.

■