STAT491

Summer Research Project

2006– 2007

An Investigation of IQ Movement In Early Childhood

Lisa Henley

Department of Mathematics and Statistics
University of Canterbury

An Investigation of IQ Movement in Early Childhood

Lisa Henley

February 2007

Abstract

The aim of this work is to investigate whether there is movement over time in early childhood IQ, and if so whether this movement differs between pre-term and full term children. Movement was indeed observed for both pre-term and full term children, however the 'type' of movement differed significantly between pre-term and full term children only for those categorised as normal. The analysis also examines the factors which may influence the movement. Various techniques are employed to contend with issues including small sample sizes, aligning independent IQ measurement scales and variable selection.

1 Introduction

A longitudinal study of a group of children born pre-term, and a control group of a similar sample size of children born full term is currently being conducted. The IQ of these children has been measured at the age of two using Bayley Mental Development Index (MDI), then at four and six using Wechsler Preschool and Primary Scale of Intelligence (WPPSI). Data for Ninety eight additional variables has been collected for each child. These variables are available for investigation as potential factors which may influence any movement in IQ over time. They are of mixed type i.e both categorical and continuous. The following subsections cover some theory relating to the methods used in the analysis. The first subsection discusses standardisation of the data. The second section discusses the Fisher Exact Test, a method for testing significance of differences in categorical counts when the sample sizes are small. Thirdly and fourthly, ordinal logistic regression and perturbation analysis were the methods used to aid in variable selection in preparation for building the Bayesian network. A Bayesian network can assist in building a parsimonious model, and is discussed in the last section of this introduction.

1.1 Standardisation of Measurement Scales

IQ values were provided for analysis for each of the children at the age of two, four and six. Typically, these continuous values are transformed to a categorical format using a customary mean, $\mu = 100$ and standard deviation, $\sigma = 15$.

Table 1: IQ categories

Range	Description	Category
$x \geq \mu + (1\sigma)$	Accelerated	1
$\mu - (1\sigma) \geq x < \mu + (1\sigma)$	Normal	2
$\mu - (2\sigma) \geq x < \mu - (1\sigma)$	Impaired	3
$x < \mu - (2\sigma)$	Severely Impaired	4

Categories are then as described in Table 1. It could then be expected that using this categorisation method, and providing the data is normally

Table 2: Category Proportions

Category	Percentage
Accelerated (1)	16%
Normal (2)	68%
Impaired (3)	14%
Severely Impaired (4)	2%

distributed, children could be expected to be classified approximately as described in Table 2:

It quickly becomes apparent that this is not the case in the study data, and examination revealed μ significantly < 100 and σ significantly different to 15 for each age group (two, four and six) for both the control and study (pre-term) group.

As μ is lower across all age groups and over both the control and pre-term group, it may be possible that the testing was more rigorous than 'average'. Additionally, concern was expressed regarding the use of a different IQ test at age two, than the test used at four and six. The approach employed to remove the test and tester effect was to use the control μ and σ *for each* age group to categorise all the data *for that particular* age group. This then resulted in control group data distributed approximately as outlined in Table 2. Details of μ and σ for each group, and distribution checks relating to this process can be found in the Analysis Section.

1.2 Fisher Exact Test

Contingency tables are often used to display how different treatments affect outcome. This approach was used to examine if IQ category has changed over time from age two through four to six. In this case the "treatment" is being pre-term or full term, and the "outcome" is the IQ category each child attains at each age point. There are three tables for this analysis which can be found in the analysis section. The usual approach to contingency tables is to apply the χ^2 statistic to each of the cells in the table. [8] However this test is not suitable when any of the "expected values" of the cells in the table is below ten. [3] In this case the Fisher Exact Test is more appropriate. Its null hypothesis is that treatments do not affect outcomes.

Table 3: Example Contingency Table

	Outcome X	Outcome Y	Total
Treatment A	a	b	a+b
Treatment B	c	d	c+d
	a+c	b+d	n

Using the Exact method we take a particular contingency table found as in Table 3, where the totals across the rows and columns are called the *marginal totals.* Fisher showed that the probability of obtaining any such set of data was given by the hypergeometric distribution shown in Equation 1.

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} \tag{1}$$

This formula gives the exact probability of observing this particular arrangement of the data, assuming the given marginal totals, with the null hypothesis that the odds ratio between Outcome X and Outcome Y among Treatment A and Treatment B equals to 1. That is to say that Outcome X and Outcome Y is equally likely given Treatment A or Treatment B. For example, the null hypothesis in this case may be that a child is equal likely to be classified as Category 1 irrespective of whether they are born pre-term or full term.

Fisher showed that we could only deal with cases where the marginal totals are the same as in the observed table. In addition, in order to calculate the probability of observing our data if the null hypothesis is true, we only sum the probabilities of tables as unusual or more unusual than our table. For example, if $a = 10$ and $b = 2$, with $a + b = 12$, there are only two other "more unusual" tables possible where $a = 11$ and $b = 1$, and where $a = 12$ and $b = 0$. Both cases are "more unusual" than the observed data, and in both cases $a + b$ is still equal to the marginal total of 12. Details of the Fisher Exact Tests and respective p-values for this study can be found in the Analysis section.

1.3 Ordinal Logistic Regression

Binary logistic regression methods apply when we have a categorical dependant variable of the simplest form - dichotomous. When there are more than two categories multinomial logistic regression can be used, or ideally, if the categories can be ordered then ordinal Logistic Regression is the preferred choice. The most well known of the ordinal logistic regression methods is called the proportional odds model.

The basic idea underlying the proportional odds model is re-expressing the categorical variable in terms of a number of binary variables based on internal cut-points in the ordinal scale. For example, if y is a variable on a K-point scale, we can define the corresponding binary variables, $y_c^*, c = 1, 2, 3...K-1$ by $y_c^* = 1$ if $y > c$ and $y_c^* = 0$ if $y \leq c$ [2].

If one has a set of explanatory variables, $x_j, j = 1, ...p$, then in this example there are $K-1$ binary logistic models corresponding to regressing each of the y_c^*'s separately against the x's. The proportional odds model assumes that the true coefficients or β-values are the same in all $K-1$ models, so that the only difference in models are the intercept terms. This means that the estimates from the binary models can be pooled to provide just one set of estimates.

If $\pi_1(\mathbf{x})...\pi_K(\mathbf{x})$ represents response probabilities at each of the values for a set of explanatory variables, the proportional odds model can be found as shown in Equation 2 through 4. [5]

$$F_{\mathbf{x}}(c) = p(Y \leq c|\mathbf{x}) = \pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) + ...\pi_c(\mathbf{x}), c = 1....K-1 \tag{2}$$

Cumulative logits are then formed:

$$L_c = logit[F_{\mathbf{x}}(c)] = log(\frac{F_{\mathbf{x}}(c)}{1 - F_{\mathbf{x}}(c)}) \tag{3}$$

The proportional odds model can then be expressed as:

$$L_c = \alpha_c + \beta_c'\mathbf{x}, c = 1....K-1 \tag{4}$$

The critical assumption of this kind of model is that the slopes (coefficients) are constant across all the equations. This is called the *parallel regression assumption.*

Computationally, the estimates of the equations are calculated iteratively using Maximum Likelihood. An arbitrary value (usually 0) is chosen for the

coefficient and then the log-likelihood for the equation is calculated as shown in Equation 5.

$$\ell(B) = \sum_{i=1}^{n} y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)] \tag{5}$$

The coefficients are varied and then the log-likelihood recalculated until a maximum is reached. The coefficients are then the maximum likelihood estimates for α and β.

The log-likelihood can be used to calculate the Akaike's Information Criterion (AIC). The AIC is defined as Equation 6, where K is the number of parameters in the model, i.e. the number of variables plus the intercept. The AIC penalizes for the addition of parameters, and thus selects a model that fits well but has a minimum number of parameters i.e. the parsimonious model. [6].

In this research a combination of ordinal logistic regression and stepwise AIC were used to aid in variable selection. Stepwise AIC involves both forward and backward selection, adding and removing variables until the most parsimonious model is found.

$$AIC = -2\ell(B) + 2K \tag{6}$$

1.4 Perturbation Analysis

Perturbation analysis can be used to assess the impact (on a model) of small random changes (perturbations) to variables. This is a useful tool in assessing collinearity, as multicollinearity can lead to instability in parameter estimates. The R tool [4] for this type of analysis works by adding small random "perturbations" to selected continuous independent variables. By default, these pertubations are $\mathcal{N}(0, 1)$. The model parameters are then re-estimated. This process is repeated n times (default is 100) after which a summary of the means, standard deviations, minimum and maximum of each of the estimates is produced. If collinearity is a problem, the estimates will be unstable and there will be a lot of variation.

The Perturbation Analysis tool in R can be used with categorical variables. Categorical variables are reclassified according to a table of reclassification probabilities. Random numbers determine to which category each case is reclassified at each iteration. The reclassification probabilities are

specified to make reclassification to the same category highly likely, e.g. specifying pcnt=95 would ensure reclassification would be to the same category approximately 95% of the time.

1.5 Bayesian Networks

Inferential statistics attempt to make valid predictions based on a sample of all possible observations. However sample variability means that estimates of statistics are subject to error. Increasing sample size or taking more samples can increase confidence in estimates. Alternatively, introducing prior knowledge is sometimes an appropriate response. When this knowledge is available it is possible to use Baye's Theorem (Equation 7) where we can update our belief in hypothesis H given the additional evidence E (data), and the background context c (prior knowledge).

$$P(H|E,c) = \frac{P(H|c)P(E|H,c)}{P(E|c)} \tag{7}$$

The left-hand term, $P(H|E,c)$ is known as the "posterior probability," or the probability of our hypothesis, H after considering the effect of E, the additional evidence on c, our prior knowledge. The term $P(H|c)$ is called the 'prior probability of H given c alone'. The term $P(E|H,c)$ is called the 'likelihood' and gives the probability of the evidence assuming the hypothesis H and the background information c is true. Finally, the last term $P(E|c)$ is independent of H and can be regarded as a scaling factor, which is ultimately related to the size of the sample. As the sample size increases, the relative weighting between prior knowledge and current evidence becomes more heavily weighted in favour of the current evidence.

Conditional probability is a very useful concept. There are many examples where the probability of one event is conditional on the probability of a previous one. While the sum and product rules of probability theory can anticipate this factor of conditionality, in many cases such calculations are very complex. [7]

> Consider a domain $\mathbf{U}$ of n variables $x_1 ... x_n$. Given a subset $\mathbf{X}$ of variables x_i where $x_i \in \mathbf{U}$, if we can observe the state of every variable in $\mathbf{X}$ then this is called an instance of $\mathbf{X}$ and is denoted as

$$X = p(x_i|x_1...x_{i-1}, \xi) = p(x_i|\Pi_i, \xi)\overrightarrow{k}_X \tag{8}$$

for the observations $x_i = k_i, x_i \in U$. The joint space of **U** is the set of all instances of **U**.

$$p(X = \overrightarrow{k}_X|Y = \overrightarrow{k}_Y, \xi) \tag{9}$$

Equation 9 denotes the generalized probability density of Equation 8 given $Y = \overrightarrow{k}_Y$ for a subject with current state information ξ. $p(X|Y, \xi)$ then denotes the *Generalised Probability Density Function* for **X** given all possible observations of **Y**. The joint gpdf over **U** is the gpdf **U**.

A Bayesian network for domain U represents a joint gpdf over U. This representation consists of a set of local conditional gpdfs combined with a set of conditional independence assertions that allow the construction of a global gpdf from the local gpdfs. The chain rule of probability can be used to ascertain these values.

$$p(x_1...x_n|\xi) = \prod_{i=1}^{n} p(x_i|x_1...x_{i-1}, \xi) \tag{10}$$

One assumption imposed by Bayesian Network theory is that each variable x_i, must be a set of variables that renders x_i and $x_1, ...x_{i-1}$ conditionally independent. In this way

$$X = p(x_i|x_1...x_{i-1}, \xi) = p(x_i|\Pi_i, \xi) \tag{11}$$

A Bayesian Network Structure then encodes the assertions of conditional independence in Equation 10 above. Essentially then, a Bayesian Network Structure B_s is a directed acyclic graph such that each variable in **U** corresponds to a node in B_s, and the parents of the node corresponding to x_i are the nodes corresponding to the variables in Π_i. [7].

As an example, it might rain today, and it might rain tomorrow, what then is the probability that it will rain on both days? Rain on two consecutive days are not independent events as if it rains on one day, it is more likely

Table 4: Marginal and Joint Probabilities for rain both days

	Rain Tomorrow	No Rain Tomorrow	Marginal Probability of Rain Today
Rain Today	0.14	0.06	0.20
No Rain Today	0.16	0.64	0.80
Marginal Probability of Rain Tomorrow	0.30	0.70	

to rain the next. Solving this problem involves determining the chances that it will rain today, and then determining the chance that it will rain tomorrow conditional on the probability that it will rain today. These are known as "joint probabilities." Suppose that p(rain today) = 0.20 and p(rain tomorrow given that it rains today) = 0.70. The probability of such joint events is determined by Equation 12

$$p(E_1, E_2) = p(E_1)p(E_2|E_1) \tag{12}$$

which can also be represented as equation 13

$$p(E_2|E_1) = \frac{p(E_1, E_2)}{p(E_1)} \tag{13}$$

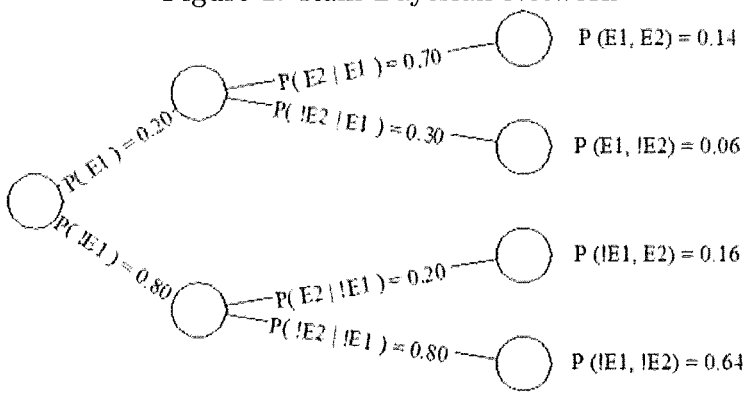
The results can be expressed as in Table 4

The joint probability of rain over both days is 0.14, but there is a lot of other information that had to be brought into the calculations to make this calculation. With only two discrete, variables, four calculations were required.

This same scenario can be expressed using a Bayesian Network Diagram as shown in Figure 1 (! is used to denote 'not'). One attraction of Bayesian Networks is the efficiency. In this case we are really only concerned with the top branch, the probability that it will rain on both days $p(E_1, E_2)$.

We can also utilize the graph both visually and algorithmically to determine which parameters are independent of each other. Instead of calculating four joint probabilities, we can use the independence of the parameters to limit our calculations to two. $p(E_1, E_2)$ and $p(E_2|E_1)$. It is obvious from the graph that the probabilities of rain on the second day having rained on the

Figure 1: Rain Bayesian Network

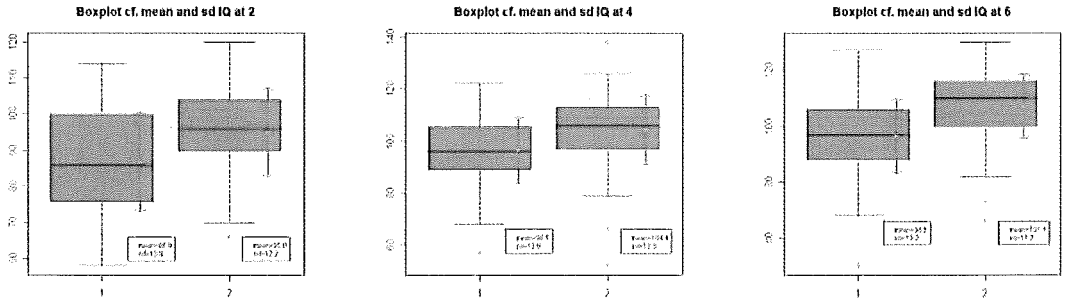


first are completely independent from the probabilities of rain on the second day having not rained on the first.

At the same time as emphasizing independence, Bayesian Networks also provide a parsimonious representation of conditionality among parametric relationships. While the probability of rain today and the probability of rain tomorrow are two discrete events (it cannot rain both today and tomorrow at the same time), there is a conditional relationship between them (if it rains today, the lingering weather means it is probably more likely to result in rain tomorrow). For this reason, the directed edges of the graph are connected to show this dependency [7].

The number of possible nets grows more than exponentially with the number of nodes. Searching for the 'best' one can be a very computationally intensive process. The R software package *Deal* [1] used to create the net in this study uses an algorithm called *greedy search with restarts.* In greedy search two models that differ only by a single arrow are compared. The arrow can be reversed, removed, or added. The network score of both nets is compared and the one which the highest score is selected.

Figure 2: Boxplot Representations of Original Data



2 Analysis

2.1 Recoding and Distribution Check

As previously described, the IQ data for each age group was recoded from a continuous variable into the categorical form described in Table 1. The mean and standard deviation of the control groups at each age point (Table 5) were used for this process. T-tests indicated statistically significant differences in the means of the control and pre-term groups at each measurement age.

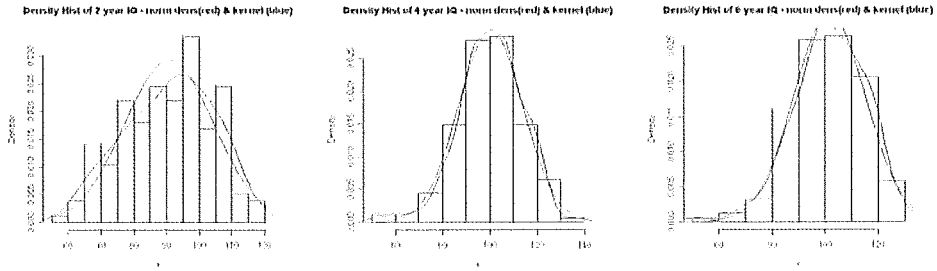
Table 5: Control Group Means and Standard Deviation

		mean	sd
Age 2	Pre-Term	86.9	13.8
	Control	95.0	12.2
Age 4	Pre-Term	96.5	12.8
	Control	104.4	13.3
Age 6	Pre-Term	96.5	13.2
	Control	107.1	11.7

Boxplots overlayed (arrows on plots) with the means and standard deviations shown in Table 5 are displayed in Figure 2.

It is always a good idea to check the data is approximately normally distributed as this is an assumption of many statistical techniques. Density histograms with the normal probability density function and kernel density functions overlayed are shown in Figure 3, and do not appear to indicate any

Figure 3: Distribution of Continuous Data



serious deviations from normality.

2.2 Contingency Tables and Fisher Exact Test

The output in this section consists of the contingency tables of the recoded data, and associated significance tests using the Fisher Exact Test described earlier.

Table 6 shows firstly the number of children in each of the categories (outlined in Table 1) at age two, and secondly, the same information shown as a percentage.

Table 6: Numbers in each category by preterm/control Age Two

	Accelerated	Normal	Impaired	Sev Impaired	Total
Pre-term	8	39	22	11	80
Control	12	52	7	4	75

	Accelerated	Normal	Impaired	Sev Impaired
Pre-term	10.0	48.8	27.5	13.8
Control	16.0	69.3	9.3	5.3

At age two there is a statistically significant difference between the categorisation of the pre-term group when compared with the control group (p-value = <0.01 for Fisher Exact Test and X-squared = 45.6027, df = 3, p-value = <0.01).

If there is no movement in IQ over time, IQ category would be expected to remain static over time. There is statistically significant evidence of

Table 7: Movement from age two to age four by preterm/control

	Accelerated	Normal	Impaired	Sev Impaired	p-value
Pre-term Age 2					
Accelerated	3	5	0	0	0.03
Normal	2	26	11	0	<0.01
Impaired	1	14	5	2	<0.01
Sev Impaired	0	3	6	2	<0.01
Control Age 2					
Accelerated	2	10	0	0	<0.01
Normal	10	39	3	0	<0.01
Impaired	0	5	1	1	<0.01
Sev Impaired	0	1	2	1	0.14

	Accelerated	Normal	Impaired	Sev Impaired
Pre-term Age 2				
Accelerated	37.5	62.5	0.0	0.0
Normal	5.1	66.7	28.2	0.0
Impaired	4.5	63.6	22.7	9.1
Sev Impaired	0.0	27.3	54.5	18.2
Control Age 2				
Accelerated	16.7	83.3	0.0	0.0
Normal	19.2	75.0	5.8	0.0
Impaired	0.0	71.4	14.3	14.3
Sev Impaired	0.0	25.0	50.0	25.0

movement between IQ at two and IQ at four for both the control and pre-term groups classified as 'Accelerated' through 'Impaired' at two years old. There was insufficient evidence of movement for control children classified as 'Severely Impaired' at two, but evidence of movement for pre-term children classified as 'Severely Impaired' at age two. Fisher Exact Test p-values are shown in Table 7 at each level.

Once again there is significant evidence of movement within the pre-term and control groups for children classified as 'Accelerated' or 'Normal' at age four, and additionally for pre-term children classified as 'Impaired' at age four. There is no evidence of movement for premature or control children

Table 8: Movement from age four to age six by preterm/control

	Accelerated	Normal	Impaired	Sev Impaired	p-value
Pre-term Age 2					
Accelerated	0	6	0	0	<0.01
Normal	4	29	14	1	<0.01
Impaired	0	5	9	8	<0.01
Sev Impaired	0	1	2	1	0.14
Control Age 2					
Accelerated	5	7	0	0	<0.01
Normal	9	46	0	0	<0.01
Impaired	0	3	2	1	0.06
Sev Impaired	0	0	0	2	1

	Accelerated	Normal	Impaired	Sev Impaired
Pre-term Age 2				
Accelerated	0.0	100.0	0.0	0.0
Normal	8.3	60.4	29.2	2.1
Impaired	0.0	22.7	40.9	36.4
Sev Impaired	0.0	25.0	50.0	25.0
Control Age 2				
Accelerated	41.7	58.3	0.0	0.0
Normal	19.2	75.0	5.8	0.0
Impaired	0.0	50.0	33.3	16.7
Sev Impaired	0.0	0.0	0.0	100.0

classified as 'Severely Impaired' at four and additionally, control children classified as 'Impaired' at four.

Is the way the movement occurs at each level ('Accelerated' to 'Severely Impaired') different between the control and pre-term babies? For example, if I am a control subject who was classified as 'Normal' at age two, is my likelihood of being classified at age four as 'Impaired' equal to that of a pre-term subject being classified as 'Impaired' at age four?

There is no evidence of a difference in 'type' of movement from age two to four between the pre-term and control children who were classified as 'Accelerated', 'Impaired' or 'Severely Impaired' at two. (Fisher exact test

p-value = 0.3473, 1 and 1 respectively). However, for children classified as 'Normal' at age two there is significant evidence of a difference in the type of movement between the pre-term and control groups (Fisher's Exact Test p-value = <0.01).

Similarly from age four to six there is no evidence of a difference in movement between the pre-term and control children who were classified as 'Accelerated', 'Impaired' or 'Severely Impaired' at four. (Fisher exact test p-values= 0.1141,0.5182,0.6). There is again significant evidence of a difference in the type of movement between the pre-term and control groups (Fisher's Exact p-value <0.01) from age four to six for children classified as 'Normal' at age four.

2.3 Bayesian Network

As previously mentioned, ninety eight additional variables were available for analysis in conjunction with the IQ data. There was a certain amount of missing data within those variables. Those records affected could not be used. Initially, a Bayesian network was built including all of the IQ data, however this gave an unsatisfactory result. Retrospective significance testing on the contingency tables showed only a difference between the pre-term and control children who were classified as 'Normal'.

Working only with the data where children were classified as 'Normal' at age 2, ordinal logistic regression was used to assist in variable selection. A factor was created consisting of eight levels with both control and pre-term children who were classified 'normal' at age two, moving to 'Accelerated' through 'Severely Impaired' from the age of two to four. That is, pre-term 'Normal' to 'Accelerated', control 'Normal' to 'Accelerated', pre-term 'Normal' to 'Normal' etc. Variables where data was available only for pre-terms were removed.

Forward selection was used as a first step to isolate variables unsuitable for use e.g. variables giving perfect separation, possibly suggesting that more data would be required for that particular variable to be useful. In combination with the regression, perturbation analysis was utilised to look for multicollinearity amongst the remaining variables. This resulted in the following variables: single parent family (0/1), oxygen at 36 weeks (0/1), proven sepsis (0/1), depression score 2 (mother), days on oxygen. The continuous variables depression score and days on oxygen were removed for simplicity in building the Bayesian network. The resulting network is shown in Figure 4

indicating move type - whether you move from the 'Normal' category at age two to the 'Accelerated', 'Normal', 'Impaired' or 'Severely Impaired' category at age four is conditionally independent of 'single parent family' and 'oxygen at 36 weeks' if you know whether a subject was pre-term or full-term, and whether or not said subject had proven sepsis or not. Deal is not designed for inference, however the conditional probabilities resulting from this network are provided (using the technique described in [1]) in Figure 5.

3 Conclusion

Although there is clearly movement among the IQ categories over time, and significant evidence of a difference between pre-term and control children when examining the movement of those children categorised as 'Normal', there were a large number of variables unable to be included in the Bayesian Network analysis as data was only available for the pre-term children. It would be interesting to undertake a similar analysis exclusively on the pre-term data to enable inclusion of these important variables.

Figure 4: Bayesian Network IQ movement age two to four

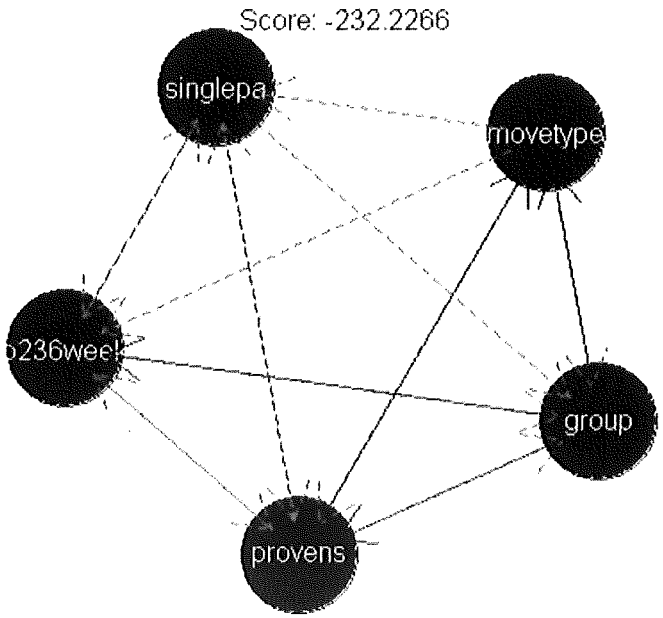


Figure 5: Conditional Probabilities

Conditional novetype|	Posterior: provens	group						
Accelerated			**Accelerated**					
	Premature	Control		Premature	Control			
No Sepsis	3.647059	12.647059	No Sepsis	0.08	0.19		1	10
Proven Sepsis	3.647059	2.647059	Proven Sepsis	0.08	0.04		1	0
Normal			**Normal**					
	Premature	Control		Premature	Control			
No Sepsis	21.647059	38.647059	No Sepsis	0.45	0.60		19	36
Proven Sepsis	6.647059	2.647059	Proven Sepsis	0.14	0.04		4	0
Impaired			**Impaired**					
	Premature	Control		Premature	Control			
No Sepsis	6.647059	5.647059	No Sepsis	0.14	0.09		4	3
Proven Sepsis	5.647059	2.647059	Proven Sepsis	0.12	0.04		3	0
	47.882354	64.8824						

Bibliography

[1] Dethlefsen C. Bøttcher, S.G., *deal: a package for learning bayesian networks*, viewed December 2006 - February 2007, http://www.jstatsoft.org/v08/i20/updates/deal.pdf.

[2] Rollin Brant, *Ordinal logistic regression*, viewed February 2007, http://stat.ubc.ca/~rollin/teach/643w04/lec/node62.html.

[3] Wikipedia The Free Encyclopedia, *Fisher's exact test*, viewed February 2007, http://en.wikipedia.org/wiki/Fisher%27s_exact_test.

[4] John Hendrickx, *Perturbation analysis*, viewed February 2007, http://rss.acs.unt.edu/Rdoc/library/perturb/html/perturb.html.

[5] Ji-Hyan Kim, *Assessing practical significance of the proportional odds assumption*, viewed February 2007, http://stat.soongsil.ac.kr/~jhkim/Publication/stat&prob_2003.pdf, 2003.

[6] Marc J. Mazerolle, *Making sense out of akaike's information criterion*, viewed February 2007, http://www.theses.ulaval.ca/2004/21842/apa.html, 2004.

[7] Daryle Niedermayer, *An introduction to bayesian networks and their contemporary applications*, viewed February 2007, http://www.niedermayer.ca/papers/bayesian/.

[8] College of Saint Benedict / Saint Johns University, *Fisher exact test*, viewed February 2007, http://www.physics.csbsju.edu/stats/exact.html.