The Application of Ordinal Regression Models in Quality of Life Scales used in Gerontology

Ranjit Lall

A thesis presented for the degree of Doctor of Philosophy

University of Sheffield

Department of General Practice and Primary Care

February 2004

Abstract

Aim of the Thesis

The area of health-related quality of life has received increasing attention particularly in gerontology. As this area grows in importance, issues such as the design and analysis of instruments that measure this multi-dimensional outcome need to be addressed. Ordinal regression models are statistical methods that can be used to analyse ordered health-related quality of life measures. However, their use is limited in the literature. The aims of this thesis are (i) to compute all ordinal regression models and compare these models with other statistical methods (such as linear regression and binary logistic regression models) and (ii) assess the use of the stereotype ordinal regression model.

Procedure

The data used to implement the regression models was from the Medical Research Council Cognitive and Function Ageing Study (MRC CFAS). In particular, two measures were chosen: the Townsend Disability Scale and the Health Status question.

Results

Linear regression models were found to summarise the ordinal data inadequately given both ordinal measures. Binary logistic regression models were only adequate for analysing ordinal quality of life scales, if one could assume that the odds ratios were the same over all the binary groupings of the ordinal scale. However, one may still encounter other problems related to multiple testing or different effects in different models. Ordinal regression models provide a more sensitive and comprehensive analysis. These methods are easily adapted to different types of ordinal quality of life data. The 'best-fit' ordinal regression model for the health status ordinal categories was the partially constrained adjacent category model. The 'best-fit' model for the Townsend Disability Scale was the fully constrained continuation ratio model.

Conclusions

This study has provided a method (based on first principles) of implementing all ordinal regression models. The comprehensive results from this thesis, suggest that ordinal regression models are indeed superior compared to other methods for analysing ordinal quality of life data. Evidence suggested that the stereotype model was of little use.

Key words: Gerontology, health-related quality of life, ordinal regression models

Declaration

The candidate, Ranjit Lall, has composed the work presented in this thesis. The work has not been submitted in any previous application of a degree.

The work (of which this thesis is a record) had been done by the candidate. All sources of information have been acknowledged.

Ranjit Lall

Acknowledgements

I would like to acknowledge with gratitude the many people who have supported me during this PhD.

First and foremost, I thank Professor Mike Campbell (University of Sheffield) and Professor Kevin Morgan (University of Loughborough) my supervisors. I owe much to them for their guidance and support in the preparation of this thesis.

My thanks also go to Stephan Walters for his input into computing models using the bootstrap technique. I am grateful to the MRC Biostatistics Unit at Cambridge for providing the dataset that was used in this thesis.

Finally, the support given by my husband, Lak and my parents has been absolutely central to this undertaking. To them and my daughter, Camani, I dedicate this thesis.

Table of Contents

List of Tables

CHAPTER 6

CHAPTER 7

CHAPTER 8

CHAPTER 1 - INTRODUCTION AND AIMS

1.1 Introduction

The main purpose of this thesis is to analyse health-related quality of life data measured on elderly respondents, using Ordinal Regression Models. Comparison of the results from these methods is made with results from other statistical methods (linear regression and logistic regression models). This aim covers two broad areas:

(i) health-related quality of life instruments in elderly respondents (in particular two measures);

(ii) ordinal regression models.

In the literature health-related quality of life in gerontology is a relatively new area and the application of Ordinal Regression Models is under-utilised when analysing ordered scales. This thesis attempts to bring together these two relatively new areas of research.

1.1.1 Quality of life in elderly people

Recent years have seen a remarkable explosion of research into quality of life in health care (often termed *health-related quality of life*). This has been particularly true in areas like gerontology, where as a result of ageing populations, elderly people have become a group of growing importance in terms of health provision and research. Ill-health and morbidity quite often dominate the health of many elderly people, and as a result it has been increasingly recognised that for most elderly people, health-related quality of life, after 75 or so years of age is more important then length of life (Cassel, 1994). This has led to a somewhat different set of priorities for medical research and health care for these people, compared to the rest of the population. In the most recent publication by the Department of Health (DOH, 2001), priorities have been outlined which aim to enhance the well being of elderly people. The immediate research priority involves exploring issues related to the dynamics of ageing and quality of life of older people. In many studies in gerontology that are being designed or are already underway to assess these issues, the different stages namely - the design of the study, choice and implementation of the quality of life instruments, the collecting, analysing and presenting of results of the quality of life data – present some challenging and relatively new aspects. This is primarily because little exists in terms of a pool of information or wealth of experience that address many aspects of health-related quality of life in elderly people.

1.1.2 Measuring health-related quality of life

(a) Single/multiple items

Generally, data on health-related quality of life are often captured on instruments or questionnaires that are either *single-item* or *multi-item*. Each question on a health-related quality of life questionnaire is an expression of words in a form of *an item*. The items on an instrument are often scaled using integers that correspond to severity. A *Global* or *single-item question* describes the quality of life of the patient, by considering all aspects of his/her health.

Many authors favour the use of *multiple-item questionnaires* as opposed to a single-item question for describing quality of life. The items can be used to either: (a) produce an average/total score, where all questions are considered to have equal weights; (b) produce an average/total score, where some questions have greater weighting depending on subjects opinion, or (c) be subdivided into groups which correspond to the different aspects or dimensions of quality of life. In the latter case, it is meaningless to produce an overall score, as quality of life is considered as a multi-dimensional concept and each group reflects the dimensions. In such circumstance, each dimension may be summarised using its own score. Such summaries are of course based on the notion that there is an underlying factor contributing to each measured item.

(b) Nominal/Ordinal Scales

In multi-item questionnaires, each item is often measured using a categorical scale, which is either *nominal* or *ordinal* in nature. Examples of a nominal scale include the items found on the Geriatric Depression Scale (Yesavage et al., 1983), where all the items in this instrument are scored nominally as 'yes'/'no'. Items on the Hospital Anxiety Depression Scale (HADS – Zigmond et al., 1983) are scored on an ordinal scale of 1-'*Not very much*', 2- '*Only a little*' and 3 –'*Hardly at all*'. The aggregated score on a multi-item instrument or the assessment made on a single-item instrument is usually captured on an *interval* or ordinal scale. For instance, in the Hospital Anxiety and Depression Scale the seven items are scored and then summed and the final score ranges from 0 to 21 on an interval scale. This score is then divided into a three-category ordinal scale: '*Normal* (<7)', '*Borderline* (8-10)' and '*Clinical depression* (11+)'.

The single-item question on health-status found on the SF-36 Health Survey (Ware et al., 1998) is rated using an ordinal scale. This question asks 'In general would you say your health is:- '*Excellent*', '*Very good*', '*Good*', '*Fair*' or '*Poor*'?'.

In nominal scales, the order of the list of the categories is unimportant. Ordinal scales usually consist of a collection of naturally ordered categories and the quantitative difference between the categories is not necessarily known. An interval scale is one that has all the characteristics of an ordinal scale and in addition the distance or difference between any two numbers on a scale have meaning.

In this thesis, attention is focused on ordinal scales used to assess the overall outcome (as opposed to item-specific outcome) using a quality of life instrument. In the case, where the ordinal outcome is based on a single-item, this outcome is termed as an *assessed variable* (Anderson, 1984). However, given that there is a continuous score derived from several items (such as in the multi-item scales) and this score has been grouped into ordinal categories, then this ordered scale is known as a *group continuous variable*.

1.1.3 Statistical methods for analysing ordinal response data

The methods used to analyse ordinal quality of life scales are rarely cited in the literature. However, in epidemiological research, the statistical methods used to analyse ordinal response data, whether assessed or grouped continuous, are frequently inappropriate (Scott et al., 1997). In general, the analysis of ordinal data is carried out by:-

a. *treating the ordinal scale as categorical*: In treating the ordinal response data as categorical, the ordinal nature of the *y*-response categories is completely ignored, which results in loss of information and considerable loss of statistical power potentially leading to incorrect inference. In addition, it is not amenable to statistical adjustment (Ananth et al., 1997; Scott et al., 1997). In testing the association of the *y*-response and the covariates, the Pearson's chi-squared test of independence or the Mantel-Haenzel test can be utilised. The Pearson's chi-squared test statistic depends on the row and the column marginal totals, but not on the order in which the rows and columns are listed. In testing independence, the χ^2-test statistic refers to the most general alternative hypothesis possible whereby cell probabilities exhibit any type of statistical independence. The χ^2-test statistic is designed to detect any type of pattern for additional parameters. In achieving this generality it sacrifices sensitivity for detecting particular patterns. Thus, the expected frequencies and the χ^2-test statistic do not change with arbitrary reordering of the columns and/or rows.

b. *Collapsing the ordinal scale into binary categories*: An ordinal outcome may be analysed using binary logistic regression. For this the response categories are treated as binary, either by collapsing the ordinal outcome into two response categories or by creating several binary categories. In either case there is loss of information in terms of the ordering of the *y*-response. Also, amalgamating the response categories can mask the true outcome, especially if there is considerable 'noise' in the data. For example, there may be 'floor' and 'ceiling' effects where subjects with very poor health who obtain the minimum scores may have no scope to register any further deterioration or improvement on the scale, and as a result, a large proportion of subjects may occur on one of the two extremes of the scale. In amalgamating the categories, these effects are either removed or accentuated. In either case, the decision to dichotomise remains arbitrary and this has to be borne in mind when the results are presented. Scott et al. (1997) showed that in the use of binary logistic regression, in which an arbitrary cut-point is selected to dichotomise the ordinal outcome, the results can lead to an estimate of the effect that is applicable only for that particular cut-point: inference outside the boundaries of that cut-point may be incorrect. Thus, binary logistic regression, according to Scott et al. (1997) does not provide an adequate summary of the data, especially if there is statistical variability in the cut-point-specific estimates of the odds ratios as well as a loss of information when the multinomial nature of the data is not accounted for. A similar conclusion was drawn by Strömberg (1996), who showed using a simulation exercise, that when changing the outcome categories on an ordinal scale (e.g. collapsing them) the effect estimate as well as the inference being drawn raised some concern.

c. *Treating the scale as interval*: The levels of the ordinal scale can be quantified and the response categories treated as coming from a continuous distribution. This, however, can cause misinterpretation of the data, in that the difference in category 6, say, and category 5 is assumed to be equivalent to the difference between category 2 and 1, given a 6-point ordinal scale (Stucki, 1996). If the intervals between consecutive points on the scale can be considered equidistant, then the use of numerical scales may be a valid analysis (Armstrong et al., 1989). For a scale such as the Hospital Anxiety and Depression Scale (HADS), however, the values of the categories correspond to different states of depression and anxiety and treating the intervals between the categories as equivalent would not be valid. Also, in many self-rating scales the data are imbalanced, e.g. one may have 'floor' or 'ceiling' effects or there may be large amounts of missing data in some categories and not others. In such cases, the choice of ordinal categories is crucial, as results will vary depending on the categories chosen (Hastie, 1989). In terms of analysing the data, statistical methods such as linear regression models (detailed by McCullagh and Nelder, 1989) or non-parametric methods such as the chi-squared test of trend, Mann-Whitney test or the Kruskall-Wallis test (Seigel, 1988) can be used to assess whether there is a trend across the levels of the response variable in relation to the covariates. Using linear regression models or the chi-squared test for trend, the degrees of freedom are reduced as a particular type of association is being examined. Thus, the methods that assess particular associations have greater power than a test like the chi-squared test, which examine a general type of association. McKelvey et al. (1975) illustrates the problem of the use of the linear regression models in the context of ordinal data. Essentially in applying linear regression models, given Y is continuous and x is a covariate, the data are expected to be normally distributed about some linear equation $Y=\alpha+\beta x$, with an error structure of zero mean and constant variance. When Y is ordinal categorical, these assumptions are generally not met. When a least squares line is fitted through the data, the error term and its variance vary for different values of x and Y. To account for these data one must assume either a non-linear model or a different error structure. Hastie (1989) states that in applying ordinary least squares regression one may produce an unbiased parameter estimate, but the corresponding estimate of variance will be biased and inconsistent. Also, Snell (1964) quite appropriately stated that for ordered categorical data, where 'floor' and 'ceiling' effects are apparent, a continuous probability model may cause problems. In the case of non-parametric methods, although non-distributional form is assumed about the data, adjustment of covariates is somewhat limited. Tests such as Cochran-Mantel-Haenszel test (Seigel, 1988) does allow for the adjustment of one covariate given another, but does not control for several covariates.

It is evident that methods used for binary or interval data cannot fully take account of the properties of ordered outcome. Statistically powerful methods, referred to as '*Ordinal Regression Models*' (Ananth et al., 1997) are the most appropriate methods for analysing ordinal data, as they take full advantage of the ordering of the *y*-response. Ordinal regression models offer some interesting analytical options: (i) these methods provide a more sensitive analysis than would be possible by arbitrarily dichotomising the outcome variable and do so without imposing unverifiable assumptions regarding the structure of the data; (ii) by modelling the dependence of an ordinal variable on a number of explanatory variables with an adjusted estimate of the effect in the form of a summary odds ratio; (iii) confounding and interaction can be assessed for all types of independent covariates: discrete categorical and continuous. Despite this, these methods have been under-utilised in biomedical and epidemiological research (Scott et al., 1997; Ananth et al., 1997; Agresti, 1999; Bender et al., 2000). As statistical software has advanced, the use of these models has also become widespread, although it is still continuously reported in the literature that some of these models are unable to be fitted using existing statistical software. However, there still remains the problem of fully understanding the models and misinterpreting the results (Bender et al., 2000). Additionally it is frequently unclear how some routine available software programs can be used to perform the calculations for these models.

There are in general seven types of Ordinal Regression Models and these have been listed in Table 1.1 together with their founders or researchers who cite the model in some depth in their literature.

Table 1.1 Ordinal Regression Models

Ordinal Regression Models	Founders/Researchers associated with Models
Polytomous	Ananth et al (1997); Lu (1999)
Proportional Odds (Cumulative Odds)	McCullagh (1980)
Unconstrained Partial Proportional Odds	Peterson and Harrell (1988, 1990)
Constrained Partial Proportional Odds	Peterson and Harrell (1988, 1990)
Adjacent Category	Agresti (1989)
Continuation Ratio	Feinberg (1980)
Stereotype	Anderson (1984)

1.1.4 The ideal method for analysing ordinal response data – varied opinions

(a) Ordinal Regression Models v. Logistic Regression Models

Despite the drawbacks of binary logistic regression and the linear regression models in analysing ordinal responses, comparison of ordinal regression models to these other statistical methods has often been cited in the literature. For instance, Scott et al. (1997) when comparing the results of the proportional odds model with the results from a series of binary logistic regressions performed at each cut-point found that the proportional odds model produced a more stable estimate of the odds ratio and the increased use of information contained in the ordinal scale resulted in an estimate with more narrow confidence limits. On the other hand, Armstrong and Sloan (1989) carried out a simulation exercise comparing conventional binary logistic models with the proportional odds using asymptotic relative efficiency (ARE). The ARE is the limit, as the sample size increases, of the ratio of the sample sizes required for the two methods in order that each achieves the same power (or equivalently the same precision) when close to the null hypothesis. This exercise showed that if the dichotomy for simple logistic regression is close to its optimal point (creating equal numbers of 'positive' and 'negative' responders), then the power gain using the proportional odds model was modest, since the relative efficiency of the simple logistic regression was between 75% and 80% depending on the number of categories used. Manor (2000) also found similarity in results when using the logistic regression model and ordinal regression models (polytomous, proportional odds, continuation ratio and the adjacent category).

(b) Ordinal Regression Models v. Linear Regression Model

The study by Lu (1999) is the only study cited in the literature that compares the ordinal regression models with the linear regression model. He found a significant difference in results when comparing the ordered logit model (polytomous model) and the linear regression model. He states that linear regression techniques fail to model the true relationship in the data and are therefore likely to underestimate the relative impact of certain explanatory variables in the response. He found that the effect of the covariates on the *y*-response differed for both types of analyses. For some covariates, opposite effect was shown for the regression analysis compared to that on the polytomous model. For other covariates using the linear regression the effect was found to be highly unlikely whereas the effect provided by the polytomous model seemed to be more what would be expected. Lu (1999) concluded that the different conclusions drawn from both sets of analyses cast some doubt on regression techniques in the context of ordinal variables. On the other hand, Walters et al. (2001) report that given a scale has more than seven categories and the distribution of the data are well spread over those categories (and there is no sparse data), then it is useful to assume that the data were generated from a continuous distribution, especially if there is reason to believe that the underlying scale is linear. In this case the usual parametric procedures such as multiple linear regression or non-parametric tests such as the Mann-Whitney can be used.

From the above there are two points to emphasis:

(i) The assessment of health-related quality of life within elderly people is a relatively new and important area of research. As a result it brings with it new challenges in terms of design and analysis of quality of life instruments.

(ii) Within epidemiological research some prefer Ordinal Regression Models when analysing ordered categories, whereas others feel alternative statistical methods are sufficient to serve their purpose. Ordinal regression models have been used in biometrical applications (e.g. Greenland 1994; Laara and Matthews, 1985) and for biomedical purposes (e.g. Armstrong and Sloan, 1989) and even in areas like ecology (Guisan et al., 2000). However, the application of Ordinal Regression Models in the context of health-related quality of life data collected in elderly people is very limited. In an extensive search in Medline (from 1966 to 2003) 117 citations were displayed which related to Ordinal Regression Models and only 3 of these were relevant to the assessment of health-related quality of life in elderly people. Similarly in the Science and Social Science Citation Indexes (from 1981 to 2002) of the 192 citations, 5 related to assessing ordinal quality of life outcomes in elderly people. Of the small number of relevant citations available in the literature, the use of ordinal regression models was either very vague, in that it was not clear whether the assumptions of the models have been checked, or an inappropriate regression model had been used.

With these two main points in mind, there has risen a need to examine the use of ordinal regression models using scales that measure health-related quality of life on elderly respondents.

1.2 Aims

For this thesis data were obtained from *the Medical Research Council Cognitive and Function Ageing Study (MRC CFAS)* and two sets of data were analysed: one with an assessed ordinal outcome and the other with a group continuous outcome. One of these datasets have been previously analysed using multivariate methods (MRC CFAS[1], 1998). The following hypotheses were apparent. The aim of this thesis was to prove/disprove these hypotheses.

(a) ***Hypothesis 1:*** Ordinal Regression Models are the most appropriate methods for analysing ordinal scales that measure health-related quality of life in elderly subjects.

This hypothesis has emerged as there are conflicting conclusions found in the literature when one compares Ordinal Regression Models with other methods. In this thesis, the comparison is made with the former methods and linear regression and binary logistic regression models. This hypothesis has also resulted because ordinal regression models in the area of quality of life within gerontology are not often used.

(b) ***Hypothesis 2:*** The *Stereotype model* is an attractive model for analysing many health-related quality of life scales where the categories are ordered.

In the literature the use of the Stereotype model is indeed very limited – in the databases searched only ten articles referred to the Stereotype model and of these, only four articles actually looked at the Stereotype model in any detail and attempted to fit it to the data in question. This model remains to be explored further. Its properties indicate that it may be an ideal model for analysing outcomes (Greenland, 1994) similar to those presented on quality of life scales.

In addressing these aims, the following issues that were considered particularly relevant to data collected on elderly subjects were also assessed:

(i) *sparse data* – Data on many elderly people, particularly the very frail can often be sparse. The issues that arise in this case are those of missing data and/or imbalance data resulting in skew distributions. This issue has been briefly addressed in the context of ordinal regression models: the analysis of data where there are large numbers of missing observations is beyond the scope of this thesis.

(ii) *First order interaction term* – Only one paper (DeMaris, 1991) in the literature cites interaction terms in the context of ordinal regression models (i.e. polytomous model). Pragmatically, interaction terms are expected to exist and therefore it is important that one addresses the modelling aspects together with the interpretation of them.

(iii) *More than one covariate* – Many researchers cite results from ordinal regression models that have been fitted using one covariate. Again, in practice, there is a need to assess how ordinal regression models behave given more than one covariate.

1.3 Format of the thesis

The thesis starts with a background on how health-related quality of life has become an important assessment when managing the health of elderly people (Chapter 2). Chapter 3 details a review of Ordinal Regression Models. The study design is outlined together with the quality of life scales and covariates used to fit the regression models in Chapter 4. The following chapters (Chapter 5 and 6) give extensive coverage of the way the models were fitted and checked for goodness-of-fit. Chapter 7 presents the results from the regression models fitted and gives a comparison of the results from the models. Chapter 8 details the discussion and conclusion of this thesis, by bringing together the statistical results and findings with emphasis on health-related quality of life instruments used in elderly people. In the latter chapter, the hypotheses stated in section 1.2 of this chapter are proved/disproved in the light of the results and the thesis is concluded with its contribution to the literature, its limitations and areas of further research.

CHAPTER 2 - BACKGROUND

2.1 Aims of this Chapter

This chapter provides the background to health-related quality of life in elderly people and how it has grown in importance. The issues addressed include: -

(i) the growth in population of elderly people and its impact;

(ii) how and why health related quality of life has become an important outcome measure in assessing the health of elderly people.

Section 2.2 provides a summary of demographic changes that have led to an ageing population. This has consequently resulted in an increasing numbers and proportion of elderly people. The incidence and prevalence of morbidity and disability inevitably increases with age, and therefore today a substantial demand placed on the National Health Service has come from the elderly sub-population.

Section 2.3 illustrates how factors such as health related quality of life have become important outcomes in elderly respondents, particularly in deciding the benefits of new and existing healthcare services and interventions. This sections also details what health-related quality of life means to elderly people. Finally a brief outline is given of the types of health-related quality of life instruments available and the ones used in this thesis.

2.2 Demography changes and its implications

2.2.1 General population

In the twentieth century substantial variations in the structure of populations have occurred particularly in the developed countries. Changes in the pattern of events have produced changes in age structure, and the twentieth century era has been characterised primarily by a decrease in the proportion of children in the population and increase in both the proportion of elderly people and the median age of the population. This change has been largely due to variations in number of births (fertility), increase in life expectancy and in net migration (Grundy, 1998). With regards to the elderly population it is mainly the variations in fertility and life expectancy rates that have led to relatively large proportion of elderly people (known as '*population ageing*').

2.2.2 Elderly population

In England and Wales, the number of people aged over 64 years of age has increased by nearly 80% during the past 45 years, and is projected to rise by another 33% over the next twenty years. By 2021 there will be 12.5 million people in this age group that will comprise 20% of the total population (Pettinger, 1998). Over the period 1991 to 2031, while the total population is expected to increase by 8%, the number of people aged 60-74 will rise by 43%, those aged 75-84 by 48% and those aged 85 and over by 138% (Department of Health, 1999).

The world's elderly population (65 years of age and over) is currently growing at a rate of 2.4% per year, considerably faster than the global total population. In the developed countries as a whole, the present elderly population numbers 165 million, and is projected to expand to 257 million by the year 2025. Sweden, with 17.5% of its population aged 65 and over in 1997, has the highest proportion of elderly people of the major countries in the world. Other notably high proportions (in excess of 16%) are found in Italy, Belgium, Greece and the United Kingdom. While the proportion of elderly people in less developed countries is currently low, in many cases fertility rates are now falling. This means that in the future these countries will see increases in the relative size of the older population (Grundy, 1992).

The relative growth of the elderly population has become a global issue, and its implications have been felt particularly in areas related to healthcare.

2.2.3 Morbidity and disability

The incidence and prevalence of chronic disease and disability inevitably increase with age, and as a result has substantial impact on the health of the elderly people. For instance, Tallis (1992) reports that the incidence of major neurological and musculo-skeletal causes of disability such as stroke, Alzheimer's disease, Parkinson's disease and osteoarthritis almost exponentially increase with age. Even epileptic seizures occur more commonly in old age.

Khaw (1999) showed the projected number of people in the United Kingdom aged over 65 (for years 1996-2066) unable\unlikely to perform the activities of daily living independently. Based on 1976 prevalence estimates, the number of people unable to perform activities of daily living will rise from 1.7 million in 1996 to nearly 3.5 million in 2051. A similar pattern was seen for dementia, with the number of cases projected to double from one million in 1996 to two million in 2051. Khaw (1999) concludes in this paper, that the number of various chronic diseases and disabilities are projected to increase two to threefold over the next 30 years for the over sixty-fives. If the prevalence of disability in later life continues at the present level, by the year 2031, we shall have two million more people in Great Britain with degree of disability sufficient enough to require daily personal help (Department of Health, 1999).

2.2.4 Use of the health and the welfare services

The implications of chronic disease and disability of older age people has been most felt by the healthcare services as elderly people are the largest users of health and social services. For instance, Grundy (1996) in a survey concluded that the use of personal social services was particularly high among very old people aged 85 years and over. Also, elderly people, particularly those aged 75 years and over, have greater contact with GPs/physicians than those in other age groups and are more likely to have been in hospital as in-patients/out-patients. Very high proportions of elderly people take prescribed drugs, particularly in the USA. Studies by the National Centre of Health Statistics estimate that the number of older people residing in nursing homes will increase by 58% from 1978 to 2003 if mortality ratio remains constant (Cohn et al., 1991). Along with the increased utilisation of nursing homes, the characteristics of the population are expected to change, resulting in facilities with older and more disabled residents. NHS hospital admissions have more than tripled since the war, from 3.5 million to more than 12 million each year. A large part of the increase is attributed to elderly people (Pettinger, 1998). Length of stay in hospital also increases with age and as a consequence of both higher admission rates and longer stay. The hospital beds per person per annum were found to be six times greater among the old people in their eighties than for those in their fifties in both England and Wales (Grundy, 1983). According to Pettinger (1998), 80% of people's lifetime healthcare costs are consumed during their first six years and last three. The burden they place on health services only occurs during their last three years.

The Department of Health (DOH) and providers of health care and social support have become increasingly concerned about the implications of the changes in the elderly population and the demands placed on cost and limited healthcare resources. As a result in the last ten years or so, the DOH have focused much research on looking at issues of ageism. In the DOH National Service Framework Report (2001), published on Ageing and Age-associated disease and disability, several priorities were set and one of them included enhancing the well-being and quality of life of older people.

2.3 Health-related quality of life

An increasing important aspect of gerontological research is the development and evaluation of interventions designed to improve the health and social status of elderly people. An important mechanism by which the goals of clinicians can be targeted, and their efforts evaluated, is through the assessment of quality of life of a population. Since clinicians are increasingly being asked to justify the benefit of additional services, the measurement of quality of life is becoming increasingly important. Following this, as manpower and medical resources are not infinite, it is important that investment in healthcare delivers not only longer life, but also delivers an improved or maintained quality of life.

2.3.1 Why health-related quality of life is a useful measure in elderly people

Health-related quality of life is an important outcome in assessing the health of elderly people for the following reasons:-

(i) *assessment of overall health*: Measuring the outcome of care is essential to providing quality services at the lowest unit cost (Ebrahim et al., 1993) yet current models of outcome measurement present considerable difficulties when applied to frail older people (Lundh and Nolan, 1996). As noted in the literature, the traditional medical outcome in which cure is the desired outcome is often inappropriate for many older people and this has resulted in the use of the functional model of health as an alternative framework (Wilkin and Hughes, 1986). In this approach success is primarily based in achieving maximum levels of functioning within the activities of daily living (ADL). However, as age and health are interrelated, it has become increasingly recognised that chronic illnesses affect many aspects of the lives of older people. In addition to functional and health status, measuring other aspects such as psychological, social and economic functioning need to receive equal importance. These collectively encompass the multi-dimensional outcome measure of *health-*

related quality of life and measuring this allows one to focus on the whole individual as opposed to only some of the aspect of the individual's health. Thus, coupled with clinical and economic outcomes, research incorporating quality of life assessments aims to provide a complete picture as regards health.

(ii) *Evaluation of the efficiency/effectiveness of interventions*: It is important that medical interventions do not simply prolong life of an elderly individual, without improving the quality of his/her life. For this reason, quality of life assessments are often chosen as an outcome measure in palliative studies, where a disease may not be cured, but the length of time of no disease and symptom relief can be prolonged. In such a case, a comprehensive evaluation of quality of life is often as important as assessing the relief of symptoms.

(iii) *Estimating the needs of the elderly population*: Measuring quality of life of an elderly person can provide information about service needs which may require some type of program intervention. For example, funding deficits in social activities may point to development programs/activities that would increase social interaction.

(iv) *Aiding policy-making decisions*: Quality of life assessments determine utilities such as life expectancy, which help in deciding on trends in health care.

2.3.2 Defining health-related quality of life in elderly people

Despite the enormous increase in research activity, there is no uniformity in the definition of health-related quality of life. In a study carried out by Cohn et al. (1991), residents (older individuals), family members and staff in nursing homes were asked to define quality of life using the following domains – care, social-emotional environment, physical environment, abilities, autonomies and morale. In general individuals differed from one another in their perceptions of health-related quality of life. For instance, residents defined quality of life in terms of care given. However, staff and family members rated physical health as more important to residents' quality of life then did the residents themselves. This study emphasised the differences in the definitions of health-related quality of life perceived by older people and by others.

There are similarities/differences in many definitions of health-related quality of life (by Fletcher et al. (1992); the U.S. Institute of Medicine (1996)), most notably in the emphasis on the multi-dimensional aspects. The most common dimensions of quality of life in elderly subjects have been detailed by Arnold (1991) and include:

(i) *physical functioning and symptoms* - mobility, self-care, fatigue, nausea, disease-specific symptoms;

(ii) *emotional functioning and behavioural dysfunctioning* - depression, anxiety and life satisfaction;

(iii) *intellectual and cognitive functioning* - memory and alertness;

(iv) *social functioning and the existence of a support network* - well-being, carer strain, social support;

(v) *role performance* - ability to do the housework, pursue interests, recreation;

(vi) *life satisfaction*;

(vii) *health perception/status*;

(viii) *economic status*;

(ix) *pain;*

(x) *the ability to pursue interests and recreations (e.g. job, hobbies);*

(xi) *sexual functioning;*

(xii) *energy and vitality.*

2.3.3 Some aspects of health-related quality of life

Choice of Instrument: Clearly given the range of definitions of health-related quality of life and reasons for measuring it, no one measure will be useful for all purposes. The choice and practicality of an instrument are particularly important in the elderly as such aspects as burden placed on the respondent, the choice of respondent, the method of administration are important issues to address. Even when the correct instrument is chosen, other aspects, such as likely refusal rates and rates of missing data can be problematic. O'Mahoney et al. (1998) highlight the concerns of lack of compliance and missing/sparse data in research studies where quality of life is measured on elderly subjects.

Type of Instrument: Generally quality of life instruments used in elderly people fall into five broad areas: clinical and observer-based scales, generic questionnaires, disease-specific and site-specific questionnaires, dimension specific questionnaires and individualised measures (Fitzpatrick and Davey, 2000). In this thesis, focus has been placed on two generic instruments: one that measures health status of elderly subjects and the other that assesses the disability of elderly respondents.

(a) Health Status measure

Health status is a complex and multi-dimensional construct, and essentially the construct of health has been defined by four factors: avoiding illness, being healthy, healthy lifestyle and disease prevention (Worsley, 1990). However, there is no widespread consensus, conceptually or operationally for a definition of health. Self-rated health is usually the individual's perception and evaluation of his or her overall health. Frequently, it is a single rating measure with responses reported along a 4- or 5-point scale from '*excellent*' to '*poor*'. As most research studies use such a global indicator of self-rated health, sceptics argue that it is risky to base a body of research on responses to a single question. Despite this, self-rated health that is measured using a single-item questionnaire has been shown to be an independent predictor of mortality in older people (Idler and Benyamini, 1997). This is consistently shown in much of the literature where self-rated health is not just a single and most important predictor of mortality, but also a predictor of future health among older adults (Mossey and Shapiro, 1982; Jagger et al., 1988; Kaplan 1988; Idler and Angel, 1990). It is also the best predictor of the use of health services (Fylkesnes, 1991; Wolinsky and Arnold, 1988). Self-rated health also corresponds closely with general hospital care and old people's home care (Branch et al., 1981; Cohen et al., 1986). There are some predictors of self-rated health cited in the literature and these are listed in Table 2.1.

Although self-rated health is measured as a categorical response, it has often been collapsed into a dichotomous variable of greater than and equal to 'good' versus less than 'good' health (Power et al., 1998; Mackenbach et al., 1997). The justification of this practice has not been established. It may be that the categories of self-rated health represent an arbitrary classification of underlying continuous phenomena. Certainly some investigators have established that the border separating 'bad' from 'good' health is vague and implies continuity (Manderbacka, 1998; Mackenbach, 1994; Manor 2000).

Alternatively, other researchers suggest that the categories may represent intrinsically distinct health states, which are predicted by different factors. Some studies have suggested that there are different predictors for good and less than good health (Smith et al. 1994). Such investigators argue that health status is composed of two different types of models. There is the 'medical' model that tends to explain self-rated health in terms of hypochondriasis.

somatisation and disability (Barsky et al., 1992), and there is a 'socio-cultural' model that identifies the power of labelling, the nature of the 'sick role' and the importance of social behaviour factors over diagnosis of chronic disease (Fylkesnes et al., 1992). A critical difference between the two types of models is the recognition that health status has as much as social role as a medical definition. Worse health status is almost entirely related to the physical experience of adverse health – current symptoms, the use of medication and past medical history. In contrast, better health status is only to a limited extent concerned with absence of illness, but is overlain with socio-demographic factors such as age, martial status and employment status. In all it is a more complex and holistic construct that involves socio-economic advantage and self-image: it is much more that the simple absence of diseases status which mark worse health. Thus, health status is made up of parameters which progress continuously (medical model) as well as parameters which characterise deviations from the norm (socio-cultural model), and thus health status as argued by Smith et al (1994) cannot solely be considered on a continuum.

(b) Physical disability

The other aspect of quality of life assessed closely in this thesis is that of physical disability. The intertwining of physical, psychological and social well-being in elderly subjects makes independent measurement of physical functioning difficult. Yet such measurements are crucial to geriatric practice. Generally the measures of physical health can be separated into three categories: those that tap the construct of general physical health or absence of illness; those that measure the ability to perform basic self-care activities, sometimes called Activities of Daily Living (ADL); and those that measure, in addition, the ability to perform some of the more-complex activities that are associated with independent life sometimes called Instrumental Activities of Daily Living (IADL). Particular attention is focused here on the ADL. General-health measures have limited value in indicating the degree of independence an individual can attain despite disease or impairments. Gerontologists have therefore had considerable interest in developing measures that tap practical dimensions. Some of these measures deal solely with basic self-care or ADL activities; some deal solely with mobility; and other measures combine both elements into a physical-functioning measure. There is some consensus, however, regarding the items that indicate ADL activities: almost all scales include some combination of dressing, bathing, toileting, transfer and feeding. Data on ADL scales are most commonly collected by direct observations and the scores on these scales are usually based on the degree of independence attained for each function. These scores are often totalled to give some indication of physical impairment.

Table 2.1: Published studies with Health Status as the response outcome and the corresponding relevant risk factors

Authors	Rating of self-rated health	Risk factors
Mossey and Shapiro, 1982	5-point rating: excellent, good, fair, poor and bad	Age, sex, education, marital status, income, life satisfaction, urban/rural residence
Idler and Angel, 1990	5-point rating: excellent, very good, good, fair and poor	Age, sex, race, education, income, medical diagnosis, smoking status, alcohol consumption, inactivity measure, obesity index
Pijls, Feskens and Kromhout, 1993	4-point rating scale: healthy, rather healthy, moderately healthy and not healthy	Age, education, marital status, family history of chronic diseases, blood pressure, serum cholesterol, electo-cardiographic diagnoses, medication use, smoking history, alcohol consumption, physical activity
Mackenbach et al., 1994	5-point rating scale: very good, good, fair, sometimes good/bad, bad	Age, sex, level of education, employment status, marital status, alcohol consumption, smoking, exercise
Hays et al., 1996	4-point rating scale: excellent, good, fair, poor	Age, sex, race, income, education, marital status, smoking, alcohol consumption, presence of chronic conditions (diabetes, heart attacks, blood pressure etc.)
Manderbacka et al., 1998	5-point rating scale: excellent, good, average, poor, very poor	Age, sex, region of residence, BMI, frequency of exercise, drinking frequency, long-standing illness/disability
Kivinen et al., 1998	3-point rating scale: not health, moderately healthy, very healthy	Age, sex, education, marital status, depression, coronary heart disease

Summary

In summarising the background to the thesis, the following points are highlighted. These include the following:

(i) Historically, over the last hundred years or so, there has been a substantial degree of change in the total population throughout the developed countries. In particular, the change has resulted in ageing populations and the relative growth of the elderly people. This in turn has been accompanied by a growth in health care demands.

(ii) As healthcare provisions and costs are limited, policy makers and providers have become increasingly concerned about the implications of the changes in population trends and the demands placed on healthcare resources.

(iii) Health-related quality of life has been seen as an important measure that provides an overall assessment of health and is also an important factor in deciding the benefits and needs of new and existing healthcare services and interventions.

(iv) As a result of this, the Department of Health have actively been promoting research in gerontology that focuses on assessing health-related quality of life.

(v) Instruments that assess health-related quality of life fall into two main categories: single item scales and multi-item scales. In this thesis, attention is focused on:

(a) a single item – a Health Status question, and
(b) a multi-item measure – a physical disability scale.

CHAPTER 3 - A REVIEW OF ORDINAL REGRESSION MODELS

As health-related quality of life measures become increasingly important in gerontology, the need to assess the statistical methods used to analyse such measures also grows in importance. As already mentioned (in Chapter 1: Introduction and Aims), the use of health-related quality of life measures is relatively new in gerontology. Therefore there does not exist the pool of information within this area to extensively assess the statistical methods, in particular ordinal regression models that can be used to analyse these instruments. Instead the application of ordinal regression models is more concentrated in other areas of epidemiology and social sciences.

3.1 Aims of this Chapter

(a) Ordinal Regression Models

The Ordinal Regression Models reviewed in this chapter are:

- Polytomous Model;
- Proportional Odds Model;
- Unconstrained Partial Proportional Odds Model;
- Constrained Partial Proportional Odds Model;
- Adjacent Category Model;
- Continuation Ratio Model;
- Stereotype Model.

(b) Issues that arise when analysing ordinal quality of life data in elderly people

This chapter is devoted to reviewing the literature related to the use of ordinal regression models primarily within non-gerontological studies. However, from this review issues that relate to analysing ordinal scales, particularly the ones that arise when dealing with the elderly population are broadly outlined.

The issues that are particularly relevant when applying ordinal regression models to the data in this thesis include:

(i) Statistical based

- The model form for each ordinal regression model and its properties.
- How are the data summarised (i.e. summary statistics) using these models?
- How are ordinal regression models fitted in software packages (particularly *SAS* as this software package will be used)?
- What methods are used to check goodness-of-fit of ordinal regression models?
- A pragmatic view of the data has been taken in this thesis. Therefore there is a possibility that interaction terms will arise, given several covariates and this issues needs to be investigated in the context of ordinal regression models.
- Does the literature adequately cover the case of two or more covariates in relation to these models?
- As the population of study is elderly people, there is a possibility of sparse data (as already outlined in section 2.3.3). What issues arise when applying ordinal regression models in the presence of sparse data?

(ii) Medical based

- Are certain models more applicable than others under certain medical conditions?
- Similarities and differences of ordinal regression models in terms of the interpretation of the summary statistics.

Prior to considering the ordinal regression models, a brief note regarding the origin of these model is given (section 3.2). Then an outline of how the data are presented in a contingency table is illustrated (section 3.3). Following this, the statistical modelling components of an ordinal regression model and the type of models used for multinomial categories are considered in depth in sections 3.4 and 3.5 respectively. The methods used to assess goodness-of-fit are also outlined (section 3.6) and modelling aspects such as the use of several covariates and interaction terms are examined (section 3.7 and 3.8 respectively). Ordinal models in the context of sparse data are also summarised (Section 3.9). The last sections (3.10, 311 and 3.12) illustrate the use of these models given different types of data and give a comparison and the interpretation of these models as presented in the literature. Finally this chapter ends with a summary that reflects on the findings in the literature and draws out the points that are particularly relevant to the issues stated above.

3.2 A brief note regarding the origin of ordinal regression models

The use of statistical methods for ordered data can be traced back to the late 1950s (e.g. Ashford, 1959). The first attempts to assess asymmetrical problems, where an ordered categorical variable was considered as the dependent variable using multiple regression techniques were published in the later 1960s (Walker and Duncan, 1967). However, reviews on this matter did not appear until the early 1980s (McCullagh, 1980; Anderson, 1984). In a series of papers, Goodman (1979, 1985) developed log-linear models that were proposed for symmetrical problems where the association between several variables, some of which were ordinal, were studied. Since then regression models that fit ordinal data have been included in the broader category of *Generalised Linear Models* (McCullagh and Nelder, 1989).

Generally there are two main classes of models that analyse ordinal categorical data – *Log-linear* and *Logit (binary and ordinal regression)* models (Agresti, 1989). Log-linear models, which allow for the ordering of the categories, for one or more variables, are termed *association models* and they describe association patterns among the variables. With this latter approach, the cell counts are modelled in a contingency table in terms of associations among the variables, and the distinction between the *response* and the *explanatory variables* is not made. In the case of the logit models, one variable is explicitly treated as the response and the others are the covariates.

In medical research studies, it is more usual to examine the relationship of the outcome with respect to other measures, as opposed to analyse the relationships among several variables. As a result, the logit models are more appropriate for our purpose.

All the ordinal regression models reduce to the binary logistic model when the categories are collapsed into two or when only two categories are assessed. One advantage of an ordered analysis over the corresponding nominal analysis is that generally fewer parameters are needed to describe a model for the response (Greenwood and Farewell, 1988). As a result the ordinal regression models are more powerful. Also, modelling ordered categorical data is intrinsically more difficult than modelling continuous data due to the constraints on the underlying probabilities and the reduced amount of information that discrete outcomes contain.

3.3 Structure of the contingency table

For a given ordinal scale, the number of observations falling in any particular category is called a *frequency*. When summarising the results in a medical study, there is often a need to assess how these frequencies are 'behaving' with respect to another variable. These data are often presented in a form of a frequency or *contingency table*, which cross-tabulates the response with the combination of the explanatory variables. Taking the most general case, let x and Y denote two categorical variables with x having r categories, ($x_1, x_2, \ldots, x_r$) and Y having c categories ($y_1, y_2, \ldots, y_c$). These categories are often called *levels* and the combination of any given x and Y category is known as a *cell* (see Table 3.1). Let π_{ij} denote the true (but unknown) probability that (x, Y) fall in the cell in row i and column j. Then the probabilities π_{ij} form a *joint distribution* of x and Y and $\Sigma_{ij}\ \pi_{ij} = 1$.

In many experimental designs, such as clinical trials and observational studies (such as case-control, cohort and cross-sectional studies), subjects are usually entered into each experimental group, the sizes of which are decided *apriori* and therefore fixed. In this fixed sampling scheme, each level of the covariate x or the combination of the levels of the different covariates are known as *sub-populations* and the total of each sub-population, is the row *marginal total* is (n_{i+} - see Table 3.1, the number of patients on each sub-population). The data are then assumed to come from a *multinomial distribution*.

Given there are r sub-populations and c levels of Y then

$$\sum_{j=1}^{c} \pi_{1j} = 1;\ \sum_{j=1}^{c} \pi_{2j} = 1;\ \ldots\ldots\ldots\ldots \sum_{j=1}^{c} \pi_{rj} = 1$$

where $\pi_{1j} = n_{1j}/n_{1+};\ \pi_{2j} = n_{2j}/n_{2+};\ldots\ldots\ldots.;\ \pi_{rj} = n_{rj}/n_{r+}$. (3.1)

The ideal methods to assess the association of x and Y, where Y may be ordinal, are based on modelling techniques.

	y-response (Y)					
x-covariate	*(y_1)*	*(y_2)*	*(y_3)*		*(y_c)*	*Row Marginal Total*
x_1	n_{11}	n_{12}	n_{13}	...	n_{1c}	n_{1+}
x_2	n_{21}	n_{22}	n_{23}	...	n_{2c}	n_{2+}
x_i				...	n_{ic}	n_{i+}
x_r	n_{r1}	n_{r2}	n_{rj}	...	n_{rc}	n_{r+}
Column Marginal Total	n_{+1}	n_{+2}	n_{+j}	...	n_{+c}	N

Table 3.1: Illustration of the way the columns and rows are formed in a contingency (rxc) table.

3.4 Components of an ordinal regression model

(a) Generalised Linear Models

Consider the most general case of linear models, as described by McCullagh and Nelder (1989). All generalised linear models are specified by three components: a *random component*, which identifies the probability distribution of the response variable; a *systematic component*, which specifies a linear function of explanatory variables that are used as predictors; and a *link* describing the functional relationship between the systematic component and the expected value of the random component.

The *random component* consists of independent observations $Y = (Y_1, \ldots Y_N)'$ from a distribution in the natural exponential family. This family includes several important distributions as special cases, including the binomial and multinomial. The *systematic component* relates a vector $\eta = (\eta_1, \eta_2, \ldots \eta_N)'$ to a set of explanatory variables (*p* of them) through a linear model

$$\eta_i = \sum_{k=1}^{p} x_{ik} \beta_k \ . \qquad (3.2)$$

The vectors of covariates $(x_{i1}, x_{i2}, ...x_{ip})$ consist of values of the explanatory variables for $i = 1, ...N$ individuals. The $\{\beta_k\}$ are the model parameters and the $\{\eta_i\}$ are called the *linear predictors*.

In defining the *link*, let $\mu_i = E(Y_i)$, $i=1, ...N$. Then μ_i is linked to η_i by $\eta_i = F(\mu_i)$, where F is any *monotonic* (i.e. as Y increases x_{ik} increase or as Y decreases x_{ik} decreases) differentiable function. The model links expected values of observations to explanatory variables through the formula

$$F(\mu_i) = \sum_{k=1}^{p} x_{ik} \beta_k \ . \qquad (3.3)$$

The function in (3.3) gives the identity link $\eta_i = \mu_i$ specifying a linear model for the mean response. However our interest is focused on link functions that are associated with ordinal regression models.

(b) Binary Logistic Regression Model

The *logistic regression model* is used as a starting point in describing the components of the ordinal regression models.

For data that have been grouped as presented in Table 3.1, it will be convenient to introduce auxiliary random variables representing counts of responses in the various categories. Let n_i denote the number of subjects in the i^{th} group. Then consider the case where one has a categorical response variable with two categories (the simplest case of multinomial categories). The observations may be classified as a 'success' (1) or 'failure' (0). Here the distribution is *Bernoulli* for the binary random variables specifying probabilities $\Pr(Y_i=1) = \pi_i$ and $\Pr(Y_i=0) = 1 - \pi_i$ for the two outcomes. Taking model (3.2) one can write the linear model

$$F(\pi_i) = \ln\{\pi_i / (1 - \pi_i)\} = \alpha_1 + \sum_{k=1}^{p} x_{ik} \beta_k \qquad (3.4)$$

and this is known as a *linear probability model* (Agresti, 1989). Here the appropriate link is the log odds transformation, known as the *logit.*

(c) Ordinal Regression Models

The models designed for categorical response variables having more than two categories are generalisations of logit model (3.4) and are collectively known as *ordinal regression models.*

For these models, since the underlying response is categorical, a member in group i can have response which falls into one of c possible categories (j=1,.... c). We let the indicator random variable Y_{ij} equal 1 if a member in group i has response j and equal 0 otherwise, with $\sum_{j=1}^{c} Y_{ij} = 1$. We can accumulate the Y_{ij}s together to form the response vector Y_i=(Y_{i1},... Y_{ic})'. The probability of response j is π_{ij} = E(Y_{ij}), with $\sum_{j=1}^{c} \pi_{ij} = 1$. The random vector Y_i has a multinomial distribution with probability vector $\pi_i = E(Y_i) = (\pi_{i1},....\pi_{ic})'$. We assume that each individual has covariates x_{ik}. For ordinal regression models there are many choices of link functions relating the elements of π_i to the covariates. Various choices of link functions include 'cumulative' logits, the 'continuation ratio' logits, the 'adjacent category' logits and 'generalised' logits. All these link functions can be generalised to the model (3.2) by

$$F(\pi_i) = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_{jk} . \qquad (3.5)$$

3.5 Types of Ordinal Regression Models

The seven different types of ordinal regression models listed in section 3.1 are now discussed in depth. Let $\pi_{ij} = \Pr(Y_i = y_j)$ denote the probability that the response of a member in group i with characteristics x_{ik} (k=1,... p) falls into the y_j category. Using this together with (3.5), the logits of the various ordinal regression models are specified below.

3.5.1 Generalised Logit/Polytomous Model

The *Generalised Logit* or *Polytomous* model is often used to model nominal response data (Agresti, 1989), although some researchers consider this model as an ordinal regression model (Lu, 1999; Ananth et al., 1997). This model is a straightforward extension of the logistic regression model for binary response and accommodates for multinomial responses.

(a) The form of the Polytomous Model

If Y is a categorical response with c categories, there are $\binom{c}{2}$ pairs of response for which one can construct logits. Given a certain choice of c-1 of these, the rest are redundant. In the case of the Generalised Logit model, each response category is paired with a baseline or referent category, the choice of which is arbitrary. For this reason Cox and Chuang (1984) referred to this model as the 'baseline comparison' logit model. When the last category (c) is referent, the log odds in the contingency table can be represented as:

$$\ln\left(\frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_c)}\right) \qquad j = 1,\ldots c\text{-}1 \qquad (3.6)$$

Given a set of predictors, Ananth et al. (1997) described a model using (3.6) as:

$$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_c)}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_{jk} \qquad (3.7)$$

As this latter model stands it is not identifiable, since adding a fixed constant to every β will give exactly the same predicted probabilities. To identify the model, constraints need to be placed on the parameters. Thus, α_c=0 and β_{ck}=0 are fixed to allow the parameters to be identified.

(b) What are the underlying assumptions of the Polytomous Model?

There are no assumptions required for the polytomous model. However, this model does not take account of the ordering of the categories.

(c) summary statistics – odds ratio

The polytomous model consists of c-1 logit equations or functions for a given covariate. It assumes that different pairs of levels of the outcome variable are required to assess the relationship of the response and the covariates. In this model, a covariate may influence the link function strongly in one category y_1, but has little influence in category y_2.

For a given covariate x_{il}, the parameters β_{jl} corresponds to the regression coefficients for the log-odds of $(Y_i = y_j)$, relative to the referent category $(Y_i = y_c)$ and there are $(c$-1) intercept parameters α_j. Exponentiating the regression coefficients β_{jl}, the covariate x_{il} will result in the cut-point specific odds ratios comparing $(Y_i = y_j)$ versus $(Y_i = y_c)$ for unit increase in the levels of x_{il} having adjusted for all the other covariates in the model.

(d) Computation of the Polytomous Model

When fitting the polytomous model, for optimal efficiency, one should use software that fits the c-1 logits simultaneously (Agresti, 1984). Estimates of the model parameters have smaller standard errors than when binary logistic regression software fits each component equation separately because the estimators in the separate-fit approach are less efficient.

More recently the computation of the polytomous model has become quite straight forward, and most statistical software packages accommodate the fit of this model (Cox and Chuang, 1984; Greenland, 1994). Ananth et al. (1997) fit the polytomous model using *SAS* and the procedure *CATMOD* with the logit link function. Hendrickx (2000) fits the polytomous model using *SAS* and *STATA* and uses this as a basis for fitting the stereotype model.

3.5.2 Proportional Odds Model

The prime feature of the proportional odds model is that a single summary measure (in terms of an odds ratio) is used to summarise the relationship of the ordinal response and the covariates.

The reason for using a common odds ratio is that the most optimal model can be fitted, at the same time allowing for model parsimony (i.e. estimating fewer parameters) and the simplification of the interpretation of the regression parameters. The proportional odds model allows for the ordering of the response categories through the use of cumulative probabilities.

The proportional odds model is the most commonly used ordinal regression model because it provides estimates that are easily interpretable. The bulk of the literature on ordinal regression models is based on the proportional odds model. It has potentially greater power than ordinary multi-category logit models. It can be viewed as a natural model for response variables of continuous form that have been made discrete during data collection.

(a) The form of the Proportional Odds Model

The origins of the proportional odds model root from a model that was first introduced by Walker and Duncan (1967). Their model was based on cumulative probabilities and was of the form:

$$\ln\left[\frac{\Pr(Y_i \leq y_j)}{\Pr(Y_i > y_j)}\right] = \ln\left[\frac{\pi_{i1} + \pi_{i2} + \ldots\pi_{ij}}{\pi_{ij+1} + \pi_{ij+2} + \ldots\pi_{ic}}\right] \qquad (3.8)$$

where the response is treated as binary by combining the first j categories and by combining the remaining $(c\text{-}j)$ categories. The $\Pr(Y_i \leq y_j)$ denotes the probability of a response in category y_j or below for the member in group i and is known as the *cumulative probability*. The L.H.S. of (3.8) represents the log of the cumulative odds or the logit. There are $(c\text{-}1)$ of these logits, one for each possible cut point when the response is collapsed. Equation (3.8) can be expressed in model terms as:

$$\ln\left[\frac{\Pr(Y_i \leq y_j)}{\Pr(Y_i > y_j)}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_{jk} \qquad j=1\ldots.c\text{-}1 \qquad (3.9)$$

where the $\{\alpha_j\}$ are the unknown intercept parameters and the $\{\beta_{jk}\}$ are the unknown regression coefficients corresponding to x_{ik}. Models for cumulative probabilities do not use the final one, i.e. $\Pr(Y_i \leq y_c)$ since it necessarily equals 1.

McCullagh (1980) considered model (3.9) in great detail, and derived from the model the '*Proportional Odds Model*'. For this model, it is assumed that one can combine the c-1 versions of the model (3.9), corresponding to the c-1 possible cut points of the response, into a single model in which the same slope parameter β is used for each logit. This means that the effect of covariates is assumed to be the same for each cumulative probability; it does not depend on the cut point for forming the logit. As the cut-point-specific estimates are not statistically independent, the proportional odds ratio is not a simple weighted average of these values, but rather is based on maximisation of a specific likelihood function. If a naïve attempt was made to treat the strata of the proportional odds model as independent and combine them using the Mantel-Haenszel technique, the variance of the β estimate would be underestimated, producing tight confidence limits (Scott et al, 1997). Under the assumption that $\beta_{1k} = \beta_{2k} = \ldots\ldots\ \beta_{c-1k}$, model (3.9) simplifies to

$$\ln\left[\frac{\Pr(Y_i \le y_j)}{\Pr(Y_i > y_j)}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_k \qquad j=1....c\text{-}1. \qquad (3.10)$$

(b) What are the underlying assumptions of the Proportional Odds Model?

There are two assumptions of the proportional odds model:

(i) the existence of an underlying continuous variable;

(ii) homogeneity in the cut-point specific regression parameters (or known as the *proportional odds assumption*).

The existence of an underlying continuous variable

One advantage of the cumulative links models is that the parameter estimates refer to the cumulative distribution of the manifest response (or the distribution of the underlying variable) and are therefore not heavily dependent on the actual categories used. The single latent variable underlies all the items in the scale and is sufficient to explain all but the random variability that is obtained in the data.The relatively stringent proportional odds assumption may be valid in cases where the ordinal response Y_i is related to an underlying continuous variable. In fact the primary motivation of the proportional odds model was provided by the existence of an underlying continuous and unobservable variable, denoted by Z_i. The ordered categories for the response variable Y_i are related monotonically to Z_i. Y_i is the grouped continuous ordinal variable, and originates from the continuous and interval scaled underlying variable Z_i, which has been made discrete during data collection. It is sensible to make use of this assumption when constructing models for ordinal response. Then the main interest is not in the ad hoc response categories; it is in the variable Z_i and its distribution and information about Z_i is obtained via the categorical response.

Following McCullagh's (1980) notation, the proportional odds model gives the conditional probability of Y_i given x_{ik} as

$$\Pr(Y_i \leq y_j) = F(\alpha_j + x_{ik}\beta_k) \qquad j = 1, \ldots .. c \qquad (3.11)$$

where F(.) is any convenient cumulative distribution function. In model (3.11) F(.) is the logistic function.

Although Z_i is not observable, a closely related grouped version of Z_i, Y_i is observable, where

$$Y_i = y_j \text{ if } \alpha_{j-1} \leq Z_i < \alpha_j \ ; j = 1, \ldots \ldots, c. \qquad (3.12)$$

The parameters $\{\alpha_j \text{ s}\}$ are thus the division points of the latent scale, Z_i. It is difficult to interpret $\{\alpha_j\}$ parameters unless the observed response variable is directly related to a latent variable. The proportional odds assumption then implies that all observations have a common variance (scale) on the underlying continuum, and tests of the x_{ik}-Y_i association are seen as tests of location on this continuum. The distribution of Y_i is linked to x_{ik} by postulating that the conditional function of Z_i given x_{ik} is $F(Z_i + \beta x_{ik})$ and model (3.11) follows immediately. However, despite this, it is not necessary to suppose the existence of an underlying continuous variable in order to use the group continuous model. The main reason for its use is that interpretation is easier and clearer.

Proportional Odds assumption

Testing the assumption of proportional odds is often overlooked in the literature and results using the proportional odds model have often been cited without checking the assumption (Bender et al., 1997).

Owing to the stringent model assumption of constant odds over the cut-points, the proportional odds model is the wrong method to start a valid data analysis. Only if the separate binary models are validated, should one proceed and assess the adequacy of the proportional odds model. A natural way to assessing proportionality is to examine and compare the conventional logistic fits for the dichotomised responses (Brant, 1990). According to Bender et al. (1998), the first step in the model building procedure should be a graphical check of whether the logits of each dichotomised response against the covariates appear to be similar. The assumption of constant odds over the cut-points can then be more formally checked using one of the following methods.

(i) Functional Asymptotic Regression Methodology (FARM): In the literature, several strategies have been cited which examine the constant slope assumption. One of these includes the Functional Asymptotic Regression Methodology (FARM). This method makes no reference to an underlying continuous variable. Koch et al. (1985) developed a two-stage method of estimation, which uses the Wald tests and weighted least squares. In the first stage of the procedure separate maximum likelihood analyses are used to estimate each of the cumulative logits, as functions of the same set of p explanatory variables; thus each analysis is a logistic regression using a binary response variable. If the proportional odds assumption is found to hold for any explanatory variable, then in the second stage of the FARM analysis new regression coefficients that take the proportionality into account are estimated using weighted least squares. This method has one major disadvantage in that it does not allow the assumption of proportional odds to be tested for a set of variables, while constraining another set to have proportional odds. Also, it does not have as much power as a method that uses maximum likelihood. For these reasons, the use of this method has been limited in the literature.

(ii) Likelihood Ratio Test: The likelihood ratio test statistic can be used to test the global proportional odds assumption by comparing the log-likelihood from the proportional odds model in the previous section, with the log-likelihood from those models where the assumption may not hold (e.g. the unconstrained or constrained partial proportional odds models- see below). The likelihood ratio test has the most desirable statistical properties compared to its competitors, but it requires two maximizations of the likelihood functions. Also, there is sometimes the problem of numerical difficulties (divergence) in obtaining maximum likelihood estimates for the full set of parameters.

(iii) Score Test: Peterson and Harrell use Rao's efficient score statistic (Rao, 1947; 1973) to develop a test of proportional odds. Essentially, this statistic constrains some variables to have proportional odds, whilst testing others for proportionality. It was developed in the context of partial proportional odds models (see below). However, unfortunately this test has several limitations. First, zero cells for a regressor variable at an inner value (i.e. within the middle of the scale) of the outcome may produce spuriously high chi-squared values (Peterson and Harrell, 1988; 1990). A similar problem may result when data are generally sparse or when one of the values of the outcome represents only a small fraction of the total sample size. Second, this is a global test of non-proportionality and it cannot distinguish heterogeneity associated with the adjusted variable from that associated with other covariates. To minimise both these difficulties, this test might be better performed with a crude rather than an adjusted model, as long as confounding varies little over the cut-points. Thirdly, the score test is sensitive to sample size, such that large samples may produce statistically significant p-values, when in fact there is little practical difference between the cut-point specific estimates. Bender et al. (1998) also states that the p-values produced using this test are far too small in some cases, and that the use of other techniques to investigate the proportional odds assumption are required.

(iv) Smooth residual plots: Harrell (1998) and Bender and Benner (2000) have recently proposed the use of smooth residual plots to assess the adequacy of the equal slopes assumption and the linearity assumption of the explanatory variable. These plots are obtained using *S-PLUS* and other statistical software packages have yet to provide the facilities to produce these graphical aids. Also, these plots can only be produced when the covariates are continuous or ordinal categorical.

(v) Wald Tests: In any model fitted with maximum likelihood Wald tests can be calculated by dividing the estimated regression coefficients by their respective standard errors obtained from the information matrix.

(vi) Brant's Method: Brant (1990) described yet another method for assessing the proportional odds assumption and according to Peterson and Harrell (1992) this was essentially the same as the FARM analysis, which is detailed above.

(c) Other Features of the Proportional Odds Model

An appealing requirement for ordinal data is that the model should in some sense be invariant under a reversal of category order, such that the restrictions imposed by the model are unchanged if y_1 is recoded as y_c, y_2 is recoded as y_{c-1}, etc., (these ideas underlie the concept of *palindromic invariance* - see McCullagh 1980). This implies that the magnitude of the summary estimates does not depend on the direction employed in modelling the outcome, i.e. whether the cut-points are formed using increasing or decreasing levels of severity. However, the sign of the β_k parameter is changed and the $\{\alpha_j\}$ reverse sign and order. The cumulative proportional odds model is the only ordinal regression model with this property, and the other ordinal regression models lack this property of invariance.

The proportional odds model (with the logistic link function) is also invariant under the collapsability of the Y_i categories. Hence if two adjacent response categories are pooled together and the cut-point removed, or Y_i is changed by moving the cut-points, the estimates of β_k should remain essentially unchanged, although the $\{\alpha_j\}$ are affected. Furthermore, if more cut-points are added, the model remains identical to that when less cut-points are used. This invariance to the choice of response categories is a nice feature of the model, as two researchers who use different response categories in studying an association should reach similar conclusions.

(d) Summary statistic – odds ratio

From model (3.10) the parameters $\beta_1, \ldots\ldots, \beta_p$ are unknown regression coefficients and the parameters $\{\alpha_j\}$ are also unknown (j = 1, 2, …., c-1). As j increases, the $\{\alpha_j\}$ parameters increase, reflecting an increase in the logits, as additional probabilities are added into the numerator (i.e. $\alpha_1 \leq \alpha_{2\ldots} \leq \alpha_{c-1}$). These represent the baseline logits of cumulative response probabilities. The regression parameters $\{\beta_k\text{s}\}$ describe how the log odds are related to the covariates. The $\{\beta_k\text{s}\}$ can take a minus sign for the predictor term and this usually occurs if the response categories have been reversed when forming the logits or the covariate levels have been reversed in the analysis. All models of the form (3.10) describe strict stochastic ordering. Thus, taking a covariate, say x_1, and if one takes two levels $x_{i1(row2)}$ and $x_{i1(row1)}$, it follows from (3.10) that

$$\left[\ln\left(\frac{\Pr(Y_i \leq y_j / x_{i1(row2)})}{\Pr(Y_i > y_j / x_{i1(row2)})}\right)\right] - \left[\ln\left(\frac{\Pr(Y_i \leq y_j / x_{i1(row1)})}{\Pr(Y_i > y_j / x_{i1(row1)})}\right)\right] = (x_{i1(row2)} - x_{i1(row1)})\,\beta_1. \tag{3.13}$$

The L.H.S. of (3.13) is the difference of the two log odds based on the two levels of a covariate and forms the *cumulative log odds ratio*. Essentially it computes the log odds that a subject in group i falls in the response categories $\leq y_j$ as oppose to $> y_j$, given that he/she has $x_{1(row1)}$ characteristic as opposed to $x_{1(row2)}$.

If model (3.9) holds, and x_{i1} is a continuous or an ordinal categorical covariate, then assuming there is integer spacing between the level of x_{i1}, such that $x_{i1(row2)}$-$x_{i1(row1)} = 1$, there are (c-1) cumulative log odds ratio since the cumulative odds ratio for all pairs of adjacent rows are equal. In the case of x_{i1} being a categorical covariate, this does not hold, and there are then (r-1) rows and (c-1) ways of splitting the response into two parts, giving (r-1)(c-1) cumulative odds ratios.

On a similar note, for model (3.10), given x_{i1} is continuous or ordinal, there is only one common cumulative odds ratio and model (3.10) implies uniform association. In the case of x_{i1} being a categorical covariate, there are (r-1) cumulative log odds ratios. Thus the regression parameter β_1 can be interpreted as the cumulative log odds ratio for the Y and x_{i1} association, controlling for the remaining explanatory variables, and e^{β_1} is the adjusted cumulative odds ratio.

(e) Computation of the Proportional Odds Model

Generally the proportional odds model has always been well accommodated in statistical software packages. The *LOGISTIC* and *CATMOD* procedures in *SAS* appear to be the ones that are employed most frequently for fitting this model (Bender and Benner, 2000; Scott et al., 1997; Lee et al., 1992; Peterson and Harrell, 1990; Agresti, 1989; Armstrong and Sloan, 1989; Engel, 1988). The *LOGISTIC* procedure provides the estimates of the regression coefficient and its standard error, Walds chi-squared statistic and a p-value. Bender and Benner (2000) mention the use of *S-PLUS* to fit the proportional odds model. Cox and Chuang (1984) used *BMDP3R* to fit this model.

Testing the assumption of a constant slope

The procedure *PROC LOGISTIC* in *SAS* provides a global score test of heterogeneity given the regression coefficients (Ananth et al., 1997).

3.5.3 Partial Proportional Odds Models

The proportional odds model and the partial proportional odds models are collectively termed *cumulative logit models*.

In practice, it is often difficult to find data for which a proportional odds model is a plausible description (Ananth et al., 1997). There is therefore a need for a model that allows for this assumption to be relaxed, with some explanatory variables that satisfy the assumptions of proportional odds, and others not. Thus, the primary reason for the formulation of the 'partial proportional odds models' was to relax the stringent assumption of constant odds ratio over all the cut-points for a given covariate. The assumption that a constant slope model holds, when in fact, for a given variable a constant log-odds ratio is not representative of all the log-odds ratios over the cut-points, can lead to the formulation of an incorrect model.

The partial proportional odds models were initiated by the work of Peterson and Harrell (1988, 1990) and in general there are two types of partial proportional odds models: *the Unconstrained Partial Proportional odds model* - allows some variables to have constant slope and others to vary by cut-points, and the *Constrained Partial Proportional odds model* - for those variables allowed to vary by cut-point, if a certain relationship appears to exist (e.g. there may be a linear trend in the odds ratios), constraints are placed on the parameters such that this relationship is taken into account.

(a) The form of the Unconstrained Partial Proportional Odds Model

An unconstrained partial proportional odds model takes the form:

$$\ln\left[\frac{\Pr(Y_i \le y_j)}{\Pr(Y_i > y_j)}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_k + \sum_{k=1}^{q} T_{ik}\gamma_{jk} \quad j=1....c\text{-}1. \qquad (3.14)$$

Here the x_{ik} are the values of an individual in group i with the full set of p explanatory variables, β_k are the regression coefficients associated with the p variables in x_{ik}. T_{ik} are the q covariates, such that $q \le p$ and contains the values of an individual in group i on that subset of the p explanatory variables for which the proportional odds assumption is either not assumed or is to be tested and γ_{jk} are the regression coefficients associated with the q variables in T_{ik}, so that $T_{ik}\gamma_{jk}$ is an increment associated only with the j^{th} cumulative logit and γ_{1k}=0. If γ_{jk}=0 for all j, then this model reduces to the proportional odds model. Thus a simultaneous test of the proportional odds model assumption for the q variables in T_{ik} is a test of the null hypothesis that γ_{jk}=0 for all j=2,... c-1. Since γ_{1k}=0 the model uses only $\alpha_1 + x_{ik}\beta_k$ to estimate the odds ratio associated with the dichotomisation of the y-response categories into the first category versus the rest of the categories, where the estimation of the odds ratios associated with the remaining cumulative probabilities involve incrementing $\alpha_j + x_{ik}\beta_k$ by $T_{ik}\gamma_{jk}$ (j = 2, . . , c-1).

(b) The form of the Constrained Partial Proportional Odds Model

Given the relationship of a covariate and the response is represented with non-proportional odds, then for the individual cut-point specific odds ratios, often a certain type of trend may be anticipated, e.g. a linear trend may be expected. In such a case, a constraint can be placed on the parameters in the model, so that the trend is taken into account. When the constraints are incorporated into the unconstrained partial proportional odds model, this model takes on the form:-

$$\ln\left[\frac{\Pr(Y_i \le y_j)}{\Pr(Y_i > y_j)}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_k + \sum_{k=1}^{q} T_{ik}\gamma_k\Gamma_j \qquad j=1....c-1. \qquad (3.15)$$

Here the Γ_j are fixed pre-specified scalars and Γ_1 =0. The new parameters γ_k are not subscripted by j and there are q of these. Although γ_k are not dependent on j, it is multiplied by the fixed constant scalar Γ_j in the calculation of the j^{th} cumulative logit.

(c) What are the underlying assumptions of the Partial Proportional Odds Models?

The assumptions of the partial proportional odds model are as for the proportional odds model, where proportional odds exist.

Proportional Odds assumption

The likelihood ratio and score tests, as mentioned above in section 3.5.2 (b) can be used to test for the proportional odds assumption in these models. The method used is based on partitioning the parameter space θ into (ψ, λ), then the score statistic tests hypothesis about ψ

while letting λ contain the nuisance parameters for which maximum likelihood estimates are obtained. In model (3.14) the global proportional odds assumption that $\gamma_{jk}=0$ (j=2,... c-1 since $\gamma_{1k}=0$ and k=1... p) is tested with a $p((c-1)-1)=p(c-2)$ degrees of freedom score statistic by letting ψ contain the γ_{jk} parameters and λ contain the α_j and β_k parameters. The proportional odds assumption for the x_{il} covariate is tested with a (c-2)-df score statistic by letting λ contain the α_j and β_k as before, but $\psi = (\gamma_{21}, \gamma_{31},.... \gamma_{(c-1)l})$. In model (3.15) a global test of proportional odds for q covariates against constrained non-proportional odds is a test of $\gamma_1 = \gamma_2 = \gamma_3 =... \gamma_p =0$ and thus has p-df. Tests of proportional odds for each variable separately against these same alternatives have one degree of freedom.

Several Wald tests related to proportional odds are available from the fit of the partial proportional odds models. If a variable x_{il} is fitted for unconstrained non-proportional odds, then a Wald test of association of this variable with the response variable has (c-1)-df, since the null hypothesis is H_0: $\beta_l=0$; $\gamma_{jl} = 0$. Likewise if a variable has constrained non-proportional odds, the comparable 2-df null hypothesis of no association is H_0: $\beta_l=0$; $\gamma_l = 0$. Furthermore, (c-2)-df Wald test of proportional odds can be calculated for all variables fitted for unconstrained non-proportional odds (H_0: $\gamma_{2l} = \gamma_{3l} =... \gamma_{(c-1)l} =0$). Likewise for variables fitted with a constraint, one-df Wald test of proportional odds can be computed (H_0: $\gamma_l =0$).

Goodness-of-fit tests of constraint

Given covariate x_{il}, the test of whether a single γ_l parameter fits the data as well as (c-2) γ_{jl} parameters can be obtained by using the likelihood ratio test. Here we compare the log-likelihood of unconstrained and constrained models. This gives an approximate chi-square test with (c-2)-1 df.

A score test of the goodness-of-fit of the constrained partial proportional odds model for variable x_{il}, say, can be obtained as follows. Let λ contain the parameters α_j, β_l and γ_{jl} or γ_l in a model for which a maximum likelihood fit is desired. The γ_l parameter for variable x_{il} is included among these parameters. Let ψ contain the c-2 parameters for variable x_{il}. Since both γ_l and the (c-2) γ_{jl}s are in (ψ, λ), the parameter space is over-specified, i.e. the c-2 possible departures from proportional odds for variable x_{il} are represented by c-1 parameters. The score test of the adequacy of the constraints is a test of $\gamma_{jl}=0, j=2,... c-1$ and has (c-2)-1 df, since one-df is taken up by the γ_l associated with the constraint in the model.

(d) Other Features of the Partial Proportional Odds Models

The choice of Γ_j scalars for the constrained model is determined by examining the odds ratios obtained from the unconstrained model (Ananth et al, 1997). The crucial point is that the same set of scalars be assigned to each covariate. However, Peterson and Harrell (1990) do discuss the option of having different constraints for different covariates and how model (3.15) can be adapted to allow for this.

(e) Computation of the Partial Proportional Odds Models

Peterson and Harrell (1988) detailed the computation of the partial proportional odds model using *SAS* (*PROC LOGIST*; version 5, 1986). The computation of these models even in this version of *SAS* (which has now become obsolete) was still limited, since all predictor variables had either proportional odds (one parameter to each covariate), unconstrained non-proportional odds or constrained non-proportional odds (odds in a pre-specified trend in log odds ratios). A mixture of variables with non-proportional odds and constrained non-proportional odds in the same model was not permissible.

The detail given by Peterson and Harrell (1990), regarding the test of proportional odds using a score test is also no longer available in *SAS*. In their paper, an unconstrained partial proportional odds model was implemented in *SAS*, and separate score tests were set up to assess the proportionality of different parameters.

Ananth et al. (1997) used an updated version of *SAS* with the *LOGISTIC* procedure, *SAS* supplemental library (version 5.18) to compute the constrained and unconstrained partial proportional odds models. The dataset in this paper had limitations in that only one covariate was used in the analysis, and therefore partial proportional odds models were fitted only to a single covariate and the application of a model with a mixture of variables did not arise. However, since then all the updated versions of *SAS* no longer support the supplement library and therefore the partial proportional odds models are not easily computed. Bender and Grouven (1998) computed the unconstrained partial proportional odds model using *SAS*, but there is no mention of how this was done. They do mention in the discussion section of the paper, that unfortunately, no standard software is currently available for the computation of this model. They recommend the use of separate binary logistic regressions to analyse ordinal data with non-proportional odds, at least as long as some comfortable standard software for the partial proportional odds models becomes available.

3.5.4 Adjacent Category Model

The adjacent-category model utilizes single category probabilities rather than cumulative probabilities. Agresti (1984, 1999) states that when the response categories have a natural ordering, logit models should utilise that ordering. One can incorporate the ordering directly in the construction of the logits.

(a) The form of the Adjacent Category Model

(i) Constant slope model

Ananth et al. (1997) and Agresti (1989) describe the adjacent category logistic model as modelling the ratio of the two probabilities $\Pr(Y_i=y_j)$ and $\Pr(Y_i=y_{j+1})$ where $j=1,\ldots.c-1$.

The model has the following representation:

$$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_{j+1})}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_k \qquad j = 1,\ldots\ldots\ldots\ldots c\text{-}1 \qquad (3.16)$$

Agresti (1984) states that like the cumulative logit, the constant slope adjacent-categories logit model implies stochastic orderings of the response distributions for different predictor values.

(ii) Different slopes model

Manor et al. (2000) described the adjacent-category model in a slightly different way. His version of the model was:

$$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_{j+1})}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_{jk} \qquad j = 1,\ldots\ldots\ldots\ldots c\text{-}1 \qquad (3.17)$$

(b) What are the underlying assumptions of the Adjacent Category Model?

Different versions of the adjacent category model have different assumptions. For instance model (3.16) assumes parallel slopes for the regression parameters over the cut-points, where as model (3.17) assumes that each cut-point is represented by a different slope parameter. However, when a constant slope adjacent category model has been applied, no mention is given of a test carried out to assess whether the slopes can be assumed to be parallel.

(c) Other Features of the Adjacent Category Model

Agresti (1989, 1984) describes the useful characteristic of logit models for adjacent categories, whereby these models can be equivalent to log linear models. Thus, some statistical software packages for log linear models can be used to fit an adjacent category model. Also, these models are equivalent to other models that form logits using pairs of categories (rather than groups of categories as does the cumulative logit model). For instance, the baseline category logit models contrast each response category with the final category. The adjacent category logit model having linear effect that is equivalent to the baseline-category logit model is of the form:

$$\ln\left[\frac{\Pr(Y_i = j}{\Pr(Y_i = c}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}(c-j)\beta_{jk} \qquad j = 1,\ldots\ldots\ldots c\text{-}1 \qquad (3.18)$$

The effect parameter β in the adjacent category logit model can be estimated by fitting the baseline-category logit model and replacing x_{ik} with $x_{ik}(c\text{-}j)$ in the design matrix.

(d) Summary statistics – odds ratio

Agresti (1989) states that given a cross-classification of an ordinal response and an ordinal explanatory variable x_{ik}, having assigned scores $x_{ik(row1)} < x_{ik(row2)} < \ldots x_{ik(rowr)}$, model (3.16) assumes a linear effect β_k that is the same for each adjacent pair of response categories. Ananth et al. (1997) states that the parameter β_k correspond to the log-odds of $(Y_i = y_1)$ relative to $(Y_i = y_2)$; $(Y_i = y_2)$ relative to $(Y_i = y_3)$ and so on, and there are $(c\text{-}1)$ intercept parameters α_j. Exponentiating the regression coefficient, β_k, for the k^{th} covariate x_{ik} will result in the odds ratio comparing $(Y = y_j)$ versus $(Y = y_{j+1})$ for one unit increase in the levels of x_{ik}.

(e) Computation of the Adjacent Category Model

Agesti (1989) fits the constant slope adjacent category model using *SAS* (1986) with the *PROC CATMOD* procedure. He suggests that *SPSS* and *GLIM* can also be used to fit this model. Ananth et al. (1997), states that whilst in version 5.18 of *SAS* the cumulative logit models can be fitted using the procedure of maximum likelihood estimation, the adjacent category model is fitted using weighted least squares procedure. Manor et al. (2000) fits the adjacent category model but does not detail the procedure used to compute this model.

Testing the assumption of a constant slope

Although various researchers fit the constant slope adjacent category model, no indication is given as to how the assumption of a constant slope was checked.

3.5.5 Continuation Ratio Model

Given an ordinal scale, where one is particularly interested in assessing the relative chance of a given rating, against all more favourable ones, then one would normally consider employing the *continuation ratio logits*. The continuation ratio model is best suited to circumstances where the individual categories are of particular interest. It is well-suited for failure-time data and outcomes which measure threshold points (Scott et al., 1997). Cole and Ananth (2001) used this model to analysis an ordinal scale that measured the degree of perinatal laceration. This response was measured on a five-point ordinal scale and the classification was based on the amount of tissue damage involved – an outcome that is irreversible in the sense that upon attaining level j a subject's response cannot revert to a lower level. Due to the nature of this type of outcome and that often presented in failure-time and threshold data, where individuals at a given level of an outcome must have passed through all previous levels of an outcome, the continuation ratio model would appear to be a reasonable starting point.

(a) The form of the Continuation Ratio Model

(i) Constant slope model

The form of the continuation ratio model was initially formulated by Feinberg (1980) and originated from survival time data. Various forms of the model exist, and the most common is the forward formulation model (Bender and Benner, 2000) and is written as:

$$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i > y_j)}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_k \quad \text{where } j = 1,2,\ldots\ldots c\text{-}1 \qquad (3.19)$$

Model (3.19) is described by Cole and Ananth (2001) as a *fully constrained continuation ratio* model. This model represents the probability of being in category j; as opposed to being in category greater than j. The intercept parameters denoted by α_j, j =1,2,……c-1, are the same as the cumulative logit model, but are not necessary ordered. In this model, it is assumed that the underlying odds ratios are the same and equal to β_k and the inference is based on maximum likelihood estimation. Essentially this model can be viewed as the ratio of the two conditional probabilities, $\Pr(Y_i = y_j / Y_i \geq y_j)$ and $\Pr(Y_i > y_j / Y_i \geq y_j)$, i.e. one models the odds of falling in category j as opposed to higher than category j, given that one has been in category j or higher.

By viewing the outcome as going from more severe to less severe, this model can be applied in reverse and forms the backward continuation ratios $\frac{\Pr(Y_i = y_j)}{\Pr(Y_i < y_j)}$. Because of the conditioning on adjacent cut-points, the continuation ratios, unlike the proportional odds is affected by the direction chosen for the response variable and the forward and backward ratios are not equivalent and yield different results. Thus, the continuation ratio model is not invariant under reversal of categories, unless Y is binary and one has to be careful which continuation ratio model one uses (Engel, 1988).

(ii) Different slopes model

Another form of this model is the different slopes continuation ratio model and for this the regression parameters are allowed to vary by the cut-point. This model is written as:

$$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i > y_j)}\right] = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_{jk} \qquad \text{where } j = 1,2,\ldots\ldots c\text{-}1 \quad (3.20)$$

The multinomial likelihood of this model factors into a product of the binomial likelihoods for the separate logits. The continuation ratio model has the advantage that the c-1 logits produced are asymptotically independent of one another. Thus, the estimation of the parameters for each of the *c-1* logits can be carried out separately, using the method of maximum likelihood, and the summation of the individual chi-square statistics gives the overall goodness of fit statistic for the set of the logit models. In practice, the continuation ratio model can be fitted in any statistical package that includes binary logistic regression, after suitable restructuring of the data (see below –Computation of the Continuation ratio model). As the fully constrained model is nested within the different slopes continuation ratio model, the difference in –2log-likelihood (deviance) provides a test of the validity of the assumption that the threshold-specific continuation ratios are equal, distributed as a χ^2-variate under the null with degrees of freedom equal to the difference in the number of parameters between the nested models.

Cole and Ananth (2001) describe mode (3.20) as the *unconstrained continuation ratio* model, and rather than use separate binary logistic regressions to fit the different slopes model, they attempt to fit it as a single model (see section 3.5.5. (ii)d). They also describe the *partially constrained continuation ratio model.* If homogeneity exists in some of the cut-point specific regression coefficients of the unconstrained model, then these regression coefficients can be assumed to have equal constrains.

Harrell et al. (1998) describe the *extended continuation ratio* model, for which the equal slopes assumption is released for some of the covariates. This resembles the unconstrained partial proportional odds model described above.

(b) What are the underlying assumptions of the Continuation Ratio Model?

Different forms of the continuation ratio model have different assumptions. For instance, model (3.19) assumes constant cut-point specific parameters for a covariate, whereas (3.20) allows the parameters to vary over the cut-points (imposing no assumptions).

Bender and Benner (2000) found through a simulation exercise that the bias of the regression coefficient was large if the standard constant slope continuation ratio model was applied, but the true model had unequal slopes.

(c) Summary statistics –odds ratio

In the case of the forward formulation, the odds is $\Pr(Y_i=y_j)/\Pr(Y_i>y_j)$ and this is the odds of $Y_i=y_j$ versus $Y_i > y_j$ conditional on $Y_i \geq y_j$. For a given covariate, x_{il}, the corresponding odds ratio is defined as

$$\left[\ln\left(\frac{\Pr(Y_i = y_j / x_{i1(row2)})}{\Pr(Y_i > y_j / x_{i1(row2)})}\right)\right] - \left[\ln\left(\frac{\Pr(Y_i = y_j / x_{i1(row1)})}{\Pr(Y_i > y_j / x_{i1(row1)})}\right)\right] = (x_{il(row2)}-x_{il(rowl)})\ \beta_l \ . \tag{3.21}$$

(d) Computation of the Continuation Ratio Model

The first description of a method to fit the continuation ratio model was given by Armstrong and Sloan (1989) and they showed that as a result of the independence of the c-1 logits, the continuation ratio model could be fitted using ordinary logistic regression techniques, provided the data have been re-arranged in a suitable form. In their paper *SAS* supplement library program *PROC LOGIST* (version 5; 1986) was used. However in this version of *SAS* neither the residual deviance nor a test of parallelism was provided, and as a result the authors carried out these tests by modest additional ad hoc calculations.

Scott et al. (1997) took the idea of restructuring the dataset and detailed it using examples from epidemiology. In brief, the new dataset is created by repeatedly including subsets of all observations contributing to each cut-point. Two new variables must be added to the dataset: one indicating the cut-point from which the particular subset arose and the second, a binary variable indicating a dichotomous status of the outcome at that cut-point. The continuation ratios are the obtained by performing binary logistic regressions on the restructured dataset with the new dichotomous outcome as the dependent variable, and the newly created cut-point levels as the independent categorical variable.

Cole and Ananth (2001) have recently computed the individual cut-point specific continuation ratio logits using 'patient-threshold' data. This method is slightly different to that presented by Scott et al. (1997) and Armstrong et al. (1989), in that each of the cut-point specific datasets which are based on a binary outcome, are incorporated into one dataset and this permits the individual cut-point specific continuation ratio logits to be computed, in one programming step. The advantage of the method demonstrated by Cole and Ananth (2001) is that a test of the beta parameters can be carried out, and some (or all) of the parameters can be constrained to be equal.

Testing the assumption of a constant slope

Scott et al. (1997) suggest that the assumption of a constant slope can be tested (without restructuring the data) and using *PROC LOGISTIC* in *SAS* (version 6 - using the complementary log-log link function). However, testing the assumption of homogeneity of effect over strata is limited with this method: the test performed is global, simultaneously testing all the parameter estimates for homogeneity over the cut-points.

3.5.6 Stereotype Ordinal Regression Model

The stereotype ordinal regression model was introduced by the late John Anderson (1984) as part of a general model for discrete multivariate outcomes and also arises naturally in the context of truly discrete outcomes. According to some researchers (Greenland, 1994), this model has been under-utilised and other ordinal regression model, such as the proportional odds model have been over-stated in terms of analysing ordinal data. The factors that motivate the need for the Stereotype model, as stated by Anderson (1984), include:

(i) *assessed variables*: Assessed (truly discrete) variables have a greater degree of subjectivity attached to them and they are more prone to observer error than grouped continuous variables, although there is less likely to be a uniform error structure across all categories for both types of variables.

(ii) *The ordering of the response categories*: In the regression models discussed so far (with the exception of the polytomous model), the regression parameters and consequently the logits are based on the ordering of the *y*-response categories. Therefore the proportional odds, continuation ratio and adjacent category models assess the association of *y*-response and the covariates conditional of the order that the categories occur. In this case the ordering is 'in-built' and assumed *a priori*.

In many cases one cannot be too certain about the relevance of the ordering of the response categories. The stereotype model is based on the polytomous model and therefore uses generalised logits. The polytomous model does not have the mechanism to account for the ordering of the y-response categories. Anderson (1984) took the latter model and assessed the relationship of the y-response and a given covariate. If the individual cut-point specific regression parameters were ordered (leading to the stereotype model), then one could assume that an ordered nature existed in the response categories. This is quite different to the 'ordinality' aspect of the proportional odds, continuation ratio and adjacent category models. For these latter models, the ordering is built into the formation of the logits. Therefore they are not necessarily ordered with respect to the covariates and in a sense it is not necessary to have any regressor variables. By contrast, in the stereotype model, the 'ordinality' only reveals itself through assessing the relationship of the y-response and x_{ik}.

(a) The form of the Stereotype Model

The stereotype model is a derivative of the polytomous logistic model (3.7). The polytomous model provides the best possible fit to the data, at the cost of a large number of parameters. The stereotype model aims to reduce the number of parameters by imposing constraints, without reducing the adequacy of the model.

The starting point for the stereotype model is to impose a structure on the β_{j1} , ... β_{jp} such that:

$$\beta_{j1} = -\phi_j \beta_1 \,;\; \beta_{j2} = -\phi_j \beta_2 \,;\; \ldots \quad \beta_{jp} = -\phi_j \beta_p \qquad j = 1, \ldots, c-1 \qquad (3.22)$$

Then model (3.7) becomes:

$$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_c)}\right] = \alpha_j - \sum_{k=1}^{p} x_{ik}\phi_j\beta_k \qquad (3.23)$$

As in the polytomous model, constraints are needed to make the model (3.23) identifiable. Anderson recommended setting $\phi_1 \equiv 1$, $\phi_c \equiv 0$ since $\beta_c = 0$. However, other constraints can be used. For instance, Greenland (1994) used $\phi_1 = 0$, $\phi_2 = 1$ and estimated subsequent $\{\phi_j\}$ parameters.

Alternatively, Hendrickx (2000) writes model (3.23) as:

$$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_c)}\right] = (\alpha_j - \alpha_c) + (\phi_j - \phi_c)(\beta_1 x_{i1} + ...\beta_p x_{ip}, \qquad (3.24)$$

$\alpha_c = 0$ and $\phi_1 = 1$ and $\phi_c = 0$.

As stated by Hendrickx (2000), in model (3.24), the regression parameters, $\{\beta_k$ s; $k = 1, \ldots, p\}$ no longer vary between the different levels of the outcome. The combination that best discriminates between the outcome variable is given by $\sum_{k=1}^{p} x_{ik}\beta_k$ and the distance between the outcome levels in terms of this linear predictor is given by the $\{\phi_j\}$ parameters.

(b) What are the underlying assumptions of the Stereotype Model?

There are no underlying assumptions of the Stereotype Model.

(c) Other Features of the Stereotype Model

The stereotype model is invariant under collapsing of the *Y* categories only for those categories assigned or constrained to have equal scores. The model is also invariant under coding reversal of *Y*, although such reversal will change the scores as well as the beta parameters.

There are certain factors that are specifically relevant to the stereotype ordinal regression model and these include the *dimensionality* of the model, *distinguishable* *y*-response categories and the *ordering* of the *y*-response categories with respect to the covariates.

(i) Dimensionality

The dimensionality, *d*, determined by the number of linear functions required to describe the relationship of *y* and x_{ik}. If only one linear function, $x_{ik}\beta_k$, is used, the model (3.23) is a *one-dimensional stereotype model.* One- dimensional relationships are much more common in the literature compared to the higher dimensions (Holtsbrugge and Schumacher (1991); Greenland (1994) and Ananth et al. (1997)).

In some instances, a single function of betas may not adequately fit the data, and for such cases the dimensionality may need to be increased. Replacing

$$\beta_j = -\phi_j \beta_k - \varphi_j \gamma_k \qquad j = 1,\ldots\ldots.c, \qquad (3.25)$$

into equation (3.23), where ϕ_1=1, ϕ_c=0, φ_1=0 and φ_c=1, we obtain a *two dimensional model.* Generally, a model with a lower dimension is always preferable to one of a higher dimension because of the smaller number of parameters needed. Ananth et al. (1997) briefly mentions a 2-dimensional extension, with the possibility of higher dimensions. Greenwood and Farewell (1988) pursued in fitting a two-dimensional stereotype model to their data, as they recognised that most of the covariates influenced the probability of being in the first response category but did little to distinguish between response categories 2 and 3. Some variables discriminated between response categories 2 and 3, whereas others did not. Since the response has *c*-categories, there is a choice of up to a maximum of *d* possible dimensional models, where d = min(c-1, number of covariates).

(ii) Indistinguishability

Once the dimensionality of the model has been determined, there are questions about ordering and model simplification, perhaps using distinguishability as a criterion. The concept of *indistinguishability* is described when a covariate, x_{il}, affects two response categories y_j and y_{j+1} in an identical manner (thus x_{il} is not predictive between the two categories). Anderson (1984) suggests that the possible causes of indistinguishability are: (a) intrinsic lack of distinguishability with respect to the covariates, (b) high observer error for the appropriate categories and therefore large standard errors for the ϕ_j estimates.

The hypothesis that $y=y_s$ is indistinguishable from $y=y_t$, with respect to the covariates takes on the form H_0: $\beta_s=\beta_t$. In the one dimension stereotype model, this is equivalent to asking whether there are differences among the ϕ_js (H_0: $\phi_s=\phi_t$).

Indistinguishability may be tested in higher dimensions. In the two-dimensional model, indistinguishability between $y=y_s$ and $y=y_t$ implies that $\phi_s=\phi_t$ and $\varphi_s=\varphi_t$.

In the example used by Holtbrugge and Schmacher (1991), the response categories were found to be indistinguishable with respect to the covariates and the categories amalgamated. However, Anderson (1984) does suggest that given indistinguishability, it is inadvisable to combine categories; it is often sufficient to appreciate their similarity. He uses the example that in questionnaire designs, the change in the form of words of a question corresponding to the amalgamation of two categories may not necessarily result in the amalgamation of responses in the two categories.

The recognition of indistinguishability simplifies the description of complex relationships.

(iii) Ordering of the ϕ_js.

With the questions about dimensionality and distinguishability settled, we have a regression model which is as economical as possible in the number of parameters. If the dimensionality is one, there is a further question about ordering.

Ordering in the one-dimensional stereotype model

As stated earlier, the ordering in a one-dimensional stereotype model is quite different from that of other logit models (with the exception of the polytomous model). If ordering is appropriate, the model orders the β_js (in the polytomous model) instead of ordering the odds or the link function. For the stereotype model, the 'ordering' is more directly tied to the effects of the explanatory variables and becomes a testable statement. If the dimensionality is one, ordering of the odds ratios is easily verified. If $\beta_k>0$ and the odds ratios form a decreasing sequence $e^{\phi_1\beta_k} \geq e^{\phi_2\beta_k} \geq ...e^{\phi_c\beta_k} \geq 1$, then

$$\phi_1 = 1 \geq \phi_2 .... \geq \phi_c = 0. \qquad (3.26)$$

Note, that (3.26) is not strictly ordinal as adjoining categories may be indistinguishable. If (3.26) is satisfied, then the effect of the covariates upon the first odds ratio is greater than the effect on the second, and so on.

Ordering in the two-dimensional (or more) stereotype models

When the dimensionality is greater than one, an ordering of the ϕ_js or the ψ_j s does not directly lead to ordering of the odds. Ordering in a two-dimensional model is difficult to interpret, since an ordered relationship in one covariate vector may differ from that in another or may be hidden by the effect of another.

(d) Summary statistics – odds ratio

Model (3.23) has a standard multinomial intercept with c-1 parameters for a response variable. It estimates c–2 independent scale values of $\{\phi_j\}$ for the response factor and a single beta parameter for each independent variable. In the polytomous model, the difference between β_{11} and β_{21} for a covariate x_{i1} illustrates how the log odds of the y-response for category 1/category 2 is affected by x_{i1}. In the stereotype model, the effect of the log odds for any two levels of the outcome is proportional for all the independent variables. However, the larger the difference between any two ϕ_j values, the more the log odds between the outcomes is affected by the independent variables. The β_l parameters show how the independent variable, x_{il} affects the log odds of higher versus lower scores, where 'higher' and 'lower' is defined by the ϕ_j scale. This model is most easily interpreted in terms of relative odds of different outcome levels. The scores $\{\phi_j\}$ are multiplicative on the logit scale and so modest score spacing represents large odds-ratio changes. Also one could set all the ϕ_j parameters equal to 1 if one thought the covariate effects would be nearly the same for all levels of the response.

(e) Computation of the Stereotype Model

In the literature the computation of the stereotype model has evolved a great deal over the last twenty years or so. Anderson (1984) initially computed the –2log-likelihood of this model using the quasi-Newton algorithm (Gill and Murray, 1972) as implemented in the NAG library. In this paper, it would appear that the constraints were estimated from the data, as Anderson states that the imposition of the ordered constraints may cause some difficulty. Greenwood and Farewell (1988) reported in their study that the stereotype model was not available in most computer software packages. Instead they fitted a logistic regression and the resulting β_j s were examined to decide whether the hypothesis $\beta_j = -\phi_j\beta$ would be tenable. Greenland (1994) made considerable progress in fitting the stereotype model. He suggested using either *GAUSS QUANTAL* (1992) or *STATA* (1992). Given fixed scores the stereotype model is of a generalised linear form and Greenland (1994) suggested fitting this model as a multinomial response model using a generalised linear model program or via constrained polytomous logistic regression. For estimated scores, the stereotype model is intrinsically non-linear. Greenland (1994) fits this model by a series of generalised linear models in which β_k and ϕ_j are alternatively held fixed while the other is estimated. However, the standard error estimated for the odds ratio and any inference based on these standard errors is not valid and Greenland (1994) suggested using Monte Carlo simulation to obtained the corrected p-values and confidence intervals. However, despite this, many authors continue to cite the stereotype model, but report that this model was unable to be computed due to the unavailability of software (Ananth et al., 1997; Bender et al., 1997; Bender et al., 1998; Guisan et al., 2000). The major breakthrough in the computation of the stereotype model has come more recently. Hendrickx (2000) worked along the same lines as Greenland and has devised user-friendly macros in *STATA* and *SAS*, known as *mclest*, which fit the model by estimating both the $\{\phi_j\}$ and the $\{\beta_k\}$ parameters, by hold one parameter fixed and estimating the other, and vice versa. However, his macros do not allow for indistinguishable categories or more than one-dimensional models. Also the standard errors for the parameters are conditional on the ϕ_j being known (and hence underestimated) one has to bear this in mind when basing inference on the standard errors or obtaining the odds ratios. Lunt (2001) has taken Hendrick's macros a stage further by implementing dimensionality and distinguishability into the computation of the stereotype model and hence devised some macros known as *Soreg*, which can be fitted in *SAS* and *Stata*. Lunt (2001) computed the stereotype model using the Box and Tidwell (1962) method that is described by McCullagh and Nelder (1989). Briefly this method treats the stereotype model as a non-linear function (since it contains a product of parameters). Using this technique a linear model containing a non-linear function, say $\eta_i = \alpha + \beta g(x_{ik} / \theta)$ can be estimated iteratively by fitting

$$
\begin{aligned}
&\eta_i = \alpha + \beta u_t + \gamma v_t \quad \text{where} \\
&u_t = g(x_{ik} / \theta_{t-1}) \\
&v_t = \left[\frac{\partial g}{\partial \theta}\right]_{\theta=\theta_{t-1}}
\end{aligned}
\qquad (3.27)
$$

and θ_{t-1} is the value of θ calculated from the (t-1) iteration. Then $\theta_t = \theta_{t-1} + \hat{\gamma} / \hat{\beta}$ and the process iterates to convergence. A final iteration with $v = \hat{\beta}v$ gives the variance of θ and its covariance with α and β. The advantage of this method is that the standard errors of {ϕ_js} parameters are provided. The standard errors produced by *soreg* allow for the uncertainty in estimating the {ϕ_js} parameters. Also care needs to be taken in interpreting the results of significance tests on the stereotype model produced by S*oreg*. Unlike linear models, it cannot be shown that the log likelihood follows a χ^2 distribution asymptotically. Hence the likelihood ratio χ^2 tests and corresponding p-values should be treated with care.

3.6 Goodness-of-fit and regression diagnostics

For the models fitted in this thesis, it is essential to check the goodness-of-fit to the data and also compare the fit of the models with one another.

A critical step in assessing the appropriateness of any model is to assess how well the model describes the observed data and examine its fit in relation to other models. In the case of ordinal regression models, the goodness-of-fit is assessed by: (i) examining model assumptions; (ii) examining how well the predicted values compare to the observed data and whether there are any outliers, and (iii) comparing different models.

Various tests have already been specified for examining the model assumptions (e.g. the cumulative odds - see Section 3.5.2(b)). In addition to these tests, the model assumptions of proportional odds and parallel slopes in the proportional odds and the continuation-ratio models respectively, can be checked by graphical methods (Ananth et al., 1997). Here the individual cut-point specific odds ratios are plotted against the cut-points themselves to illustrate how the odds ratios behave for the given model.

The comparison of the predicted values with the observed data usually entails two stages: computing a goodness-of-fit statistic that provides a summary measure of the errors, and examining the individual values of the errors. These two aspects are detailed below.

3.6.1 Goodness-of-fit statistics

According to Ashby et al. (1986) the goodness-of-fit for a c-category model is a natural extension of a two category one. The observed probabilities can often be compared with the predicted probabilities to assess the goodness-of-fit. For ordinal regression models, the estimated expected frequencies are values that provide the closest fit to the observed cell counts, subject to the constraints that they satisfy the model and match the observed data in certain marginal totals (Agresti, 1989). Ashby et al. (1986) derived the individual cell probabilities based on the proportional odds model. Thus, the predicted probabilities ($\hat{\pi}_{ij}$) for each category are:

$$\hat{\pi}_{ij} = \begin{cases} 1 - \hat{\pi}_{ij}; & j = 1 \\ \hat{\pi}_{ij-1} - \hat{\pi}_{ij}; & j = 2 \ldots c \end{cases} \qquad (3.28)$$

If there are only categorical covariates and hence a limited number of sub-populations, the global goodness-of-fit can be examined by well-known methods such as the Pearson's chi-squared statistic or the likelihood ratio test. These tests take on the following form respectively:

$$\chi^2 = \sum \left(\frac{y_{ij} - n_{ij}\hat{\pi}_{ij}}{\sqrt{n_{ij}\hat{\pi}_{ij}(1 - \hat{\pi}_{ij})}} \right)^2 \text{ (Pearson's chi-squared statistic)} \qquad (3.29)$$

$$G^2 = 2\sum \sum n_{ij} \log\left(\frac{n_{ij}}{n_{i+} - \hat{\pi}_{ij}} \right) \text{ (Likelihood ratio test).} \qquad (3.30)$$

Like the Pearson's chi-squared statistic, the likelihood ratio statistic is non-negative and tends to take larger values when the fit is poorer, for a given sample size. An advantage of the likelihood ratio statistic is that unlike the Pearson form, it cannot increase as the model is made more complex. However, the G^2 and the χ^2 statistics do not provide valid test of goodness-of-fit when the cell counts tend to be small (e.g. less than 5). In such cases differences in the G^2 values can still be useful in comparing complete and reduced models.

If the number of sub-populations are large and hence the number of replicated measurement is small (or the cell counts are less than 5 in frequency), for instance for sparse data or continuous predictors, these latter methods are invalid, because they require large number of replicated measurements (Bender et al., 1997). Alternatives include the Hosmer-Lemeshow (1980) statistic for binary logistic regression. For ordinal response variables, Lipsitz (1996) uses a generalisation of the Hosmer-Lemshow method to produce a goodness-of-fit test statistic (see Appendix III: section 1, for details of how this test is computed). However, the use of this test statistic has been somewhat limited, and researchers have often fallen back on the original Hosmer and Lemeshow (1980) method that appears to be sufficient for their purposes. For instance, in the study by Bender et al. (1998), the overall goodness-of-fit of the final model (proportional odds model) was investigated by means of the test proposed by Hosmer and Lemeshow (1980) based on the separate binary regressions of the cut-points using cumulative probabilities.

The goodness-of-fit of a model can also be examined using the residuals of the logit functions and applying the Wald statistic (Stokes et al., 1985).

3.6.2 Regression diagnostics

A class of statistics called 'regression diagnostics' have been proposed to examine the role of individual subjects in the model. The purpose of regression diagnostics is to aid in identifying subjects who are problematic under the current model. For binary logistic regression models two types of diagnostic statistics are often stated: measures of residual and leverage (influence).

Measures of residual include the individual components of the Pearson's chi-squared and Deviance statistics. These statistics allow one to identify poorly fitted observations. The deviance statistic, D, is given by

$$D = \sum_{i=1}^{N} d_i^2, \qquad (3.31)$$

with the individual components d_i which are used for residual diagnostic purposes and are called the 'deviance residuals'. Assuming that the fitted value from the logistic regression model are indicated by $\hat{y}_i = p_i$ (p_i is the estimated probability that y_i=1 for subject i), then if y_i=1

$d_i = \sqrt{2\,|\,In(p_i)\,|}$ and if y_i=0, then $d_i = \sqrt{2\,|\,In(1-p_i)\,|}$. The Pearson's chi-squared statistic defined in (3.29) also has individual residual components which can be used for regression diagnostics and which are summated to form the χ^2-test statistic, in 3.29.

Measures that assess the leverage essentially describe various aspects of influence. Many of these relate to the effect on certain characteristics of removing the observation from the dataset. These measures are algebraically related to an observation's leverage - its elements are from the diagonal of the so-called *hat-matrix*. Generally the greater an observation's leverage, the greater its potential influence. Influence measures for each observation include assessing the change in χ^2 (3.29) or D (3.31) goodness-of-fit statistics when the observation is deleted.

Diagnostic tools for ordinal data are very limited and undeveloped (Lipsitz, 1996; Agresti, 1999), and as a result much more effort by the user is required to find models describing the data adequately. In the literature, there is very little available on regression diagnostics for ordinal regression models. In the light of this, researchers continue to use the simpler methods of comparing the observed and expected cell probabilities/frequencies to assess any individual extreme values or 'outliers' (Holtbrügge and Schumacher, 1991). These methods are based on visual comparison of the observed and fitted values. Lesaffre (1986) has proposed extensions of tests for goodness-of-fit and logistic regression diagnostics to the polytomous logistic regression model. However, these methods are not easily calculated using available software. Begg and Grey (1984) recommend assessing fit and calculating logistic regression diagnostics using individual logistic regression models based on the cut-points of the polytomous model. One would assess the fit of the separate logistics regression models and then integrate the results, in a descriptive manner, to make a statement about the fit of the ordinal regression model. Integration of the results requires thoughtful consideration of the effects of influential and poorly fitted sub-populations on each model. In particular, the populations that are influential for only one logit, should be examined closely with due consideration to biologic issues before they are excluded from the analysis.

3.6.3 Comparison of the ordinal regression models

Models that are nested within one another, for instance the proportional odds and partial proportional odds models, different slopes and fully constrained continuation ratio models, the adjacent category (different slopes and constant slope models) and the polytomous and the stereotype models (given one covariate)), can be compared by assessing the change in the –2log-likelihood (i.e. the deviance function). However little exists to compare the model fit of all regression models. In this context, Agresti (1999) mentions two goodness-of-fit statistics: the Akaike Information Criteria (*AIC*) and the Schwartz Criterion (*SC*). These are often used to compare different models from the same data and they support model parsimony, imposing a penalty for increasing the number of parameters in the model. Smaller values represent better fit. The *AIC* takes on the form of $-2\text{Log}L+2p$ where p is the number of parameters. These statistics are rarely used to compare ordinal regression models, and according to Agresti (1999), it would be worthwhile to develop residual analysis that exploits the ordinal nature of the response, as well as to develop and evaluate indices such as the *AIC* to compare the fit of distinctly different models.

As with many statistical endeavours, there is a danger in putting too much emphasis on statistical tests, whether of effect or goodness-of-fit. Results are sensitive to sample size and test statistics merely help indicate the level of parsimony that can be achieved. Ashby et al. (1986) states that in large-scale surveys minor imperfections in a model may yield a highly significant goodness-of-fit statistic but the model can still be an informative summary of the data. The values of goodness-of-fit procedures are to see whether the model is not fitting, rather then to formally reject the model. Agresti (1999) takes this one step further and suggests that if an ordinal regression model (i.e. proportional odds or the continuation ratio) does not fit, then possible strategies include: (a) trying a different link function such as the log-log for which the response curve is non-symmetric; (b) adding additional terms, such as interactions to the linear predictor; (c) or permitting separate β_j effects for each logit.

3.7 Several covariates in an ordinal regression analysis

For both datasets used in this thesis, a pragmatic view of the data has been taken and several covariates will be used to fit the regression models.

Many studies report results for which only one covariate has been used to fit the ordinal regression models. Rarely do researchers take the pragmatic view and assess several covariates in a model. Thus issues related to several covariates have not been assessed fully in the context of ordinal regression models. One of these issues includes examining the standard error of a single and adjusted covariate.

In classic linear regression models, it is well known that when adjustment of covariates is made, very often there is an improvement in the precision of the estimates of the coefficients of the adjusting variables. In the logistic regression models, however, this does not hold, as was detailed in a paper by Robinson and Jewel (1991). In brief, these researchers found that given Y_i, x_{i1} and x_{i2} are binary variables, with Y_i as the response then when x_{i1} was fitted on its own, the standard error of the beta estimate was always less compared to when x_{i1} was fitted with adjustment for x_{i2}. The parameter estimates were compared using the Asymptotic Relative Precision (ARP), where

$$\text{ARP}(\beta_1 \text{ to } \beta^*_1) = \frac{\text{var}(\hat{\beta}_1)^{-1}}{\text{var}(\hat{\beta}^*_1)^{-1}} = \frac{\text{var}(\hat{\beta}^*_1)}{\text{var}(\hat{\beta}_1)}. \quad (3.32)$$

Here β^*_1 is the parameter estimate of fitting x_{i1} on its own and β_1 is the parameter estimate of fitting x_{i1} adjusting for x_{i2}. It was found that provided one of the following conditions exist, that is:

(i) x_{i1} and x_{i2} are strongly associated with one another or

(ii) there is a strong association between Y_i and x_{i2} given x_{i1},

then the ARP ≤ 1, implying a less precise parameter estimate for x_{i1} when allowing for x_{i2}. With regards to (ii) the stronger the association between the variables Y_i and x_{i2}, given x_{i1} that is the larger the magnitude of β_2, the poorer the precision of the estimator β_1. The way the parameters behave with respect to one another using the ARP has not been explored in ordinal logistic regression models.

3.8 Interaction terms in ordinal regression models

For each dataset in this thesis, as there are two or more covariates assessed with respect to the response, the issue of interaction terms in ordinal regression models needs to be addressed.

The examination of the interaction terms (whether first order or otherwise) is an aspect of ordinal regression models that has been rarely cited in the literature. Very few researchers actually fit or assess the interaction terms in this context. Only one paper (Demaris, 1991) gave details of the first order interaction in the polytomous model. Scott et al. (1997) mentioned in the introduction of their paper that interaction terms could be assessed in ordinal regression models, but no more was stated further. Armstrong and Sloan (1989) also mention that fitting an interaction term can proceed in exactly the same way as for logistic regression with a dichotomised response. Greenwood and Farewell (1988) fit the proportional odds model with a first order interaction term, but no explanation is given in terms of the interpretation of the interaction term. Harrell et al. (1998) states that careful fitting of a statistical model is essential so that interactions, if present, represent biologic phenomena rather than general lack of fit of the model.

3.9 Sparse and small frequency data

The data in this thesis are collected in elderly subjects, some of who may be old and frail. Therefore there is a possibility of sparse data (as this can be problematic – see section 2.3.3) and this needs to be looked at with respect to fitting ordinal regression models.

Sparse data can cause severe bias in estimators of odds ratios and poor-chi squared approximations for the goodness-of-fit statistics. According to Agresti (1999), this area is very much under researched.

Greenland (1994) in his study showed how the stereotype model is used to approximate the proportional odds models, when all but the lowest (reference) outcome level is rare. He also showed that when the scores are equal, the stereotype model approximates the reversed continuation ratio model.

3.10 Are certain models more applicable than others under certain medical conditions ?

Two broad groups can be defined in terms of the ordinal nature of the y-response categories:

(a) *group continuous* : The ordinal scale has an unobserved underlying continuum for a grouped continuous response, e.g. such as that in the HADS (Hospital Anxiety and Depression Scale), where each item is measured from 0 to 3. Here there are a total of seven questions and the scores on the items are summed to given a final score ranging from 0-21. This score is divided into a three categorical ordered scale: '*Normal*' (<7); '*Borderline*' (8-10) and '*Clinically depressed*' (11+). The categories on the scale are related to an underlying continuum, which is the final summated score.

(b) *assessed* or *judged* ordinal outcome categories. The data obtained on the Hospital Anxiety and Depression Scale are different to that, for instance, in some dimensions of the SF-36 Questionnaire. The latter questionnaire assesses the general health status of individuals and there is a question on health status that asks "In general, would you say your health is '*Excellent*', '*Very good*', '*Good*', '*Fair*', '*Poor*'?" Here the rank of the categories is known to exist on a single dimension. Although one can assume that the categories are ordered, the structure of the ordering with respect to a given explanatory variable is unknown. For this reason Anderson (1984) recognised these types of ordered categories as being truly discrete and referred to the outcome response as a judged or an assessed variable.

Greenland (1994) emphasised that the type of ordinal regression model used for analysis should depend on the way the data have been processed and generated, and that this is quite often overlooked. In the literature the proportional odds model is quite frequently used in the context of grouped continuous ordinal data, and the polytomous and stereotype models are often referred to when a truly discrete ordinal scale is present. However, the distinction of which model should be used, given a certain type of ordered scale is not clear-cut. Other ordinal regression models such as the continuation ratio and the adjacent category models are often applied to both types of ordinal categorical scales.

3.11 Similarities and differences of ordinal regression models

Constant slope models

The proportional odds model is now the most commonly used ordinal regression model because it has the convenient feature that the effect of a covariate on the *y*-response can be quantified by one regression coefficient. McCullagh (1980) and McCullagh and Nelder (1989) state that if the order of the categories can be specified with confidence a priori, models making this ordering a strong assumption (such as the proportional odds and continuation ratio models) are preferable to the more flexible logistic models such as the stereotype model, because of their simple interpretation. Once heterogeneity has been ruled out, the proportional odds model offers several advantages over binary logistic regression, including increased power and measure of effect that applies to all dichotomies of the outcome.

According to Agresti (1999) the constant slope adjacent category and the cumulative odds models both imply stochastic ordering of the response distribution for different predictor values. Agresti (1989) found that the fit of the uniform association adjacent category model to be similar to the proportional odds model.

Connection between the proportional odds and the continuation ratio models

Often in the literature, the proportional odds and the continuation ratio models are compared, and in some respects this is not surprising as the mathematical formulation for both models are very close. Both the proportional odds and the continuation ratio models are linear and additive on the logit scale. The first cut-point specific odds-ratio is identical for both models. Läärä and Matthews (1985) demonstrate the equivalence of these models. When the complementary log-log link is used with the group continuous data with known cut-points, the model becomes the same as the continuation ratio model. According to Armstrong and Sloan (1989) if the strict proportional odds model holds (i.e. cumulative log odds ratio are constant, say β^*), then the conditional log odds ratio (β_j) of the more general continuation ratio model (i.e. with different slopes) will start at $\beta_1 = \beta^*$, but then tend to 0 as j increases. Thus the proportional odds model has been proposed as an alternative to the proportional hazards, i.e. continuation-ratio model for survival data (Bennett, 1983). Essentially the same argument explains why if the proportional odds and continuation ratio models both fitted a set of data, the estimated continuation odds ratio will be less than proportional odds ratio. Greenwood and Farewell (1988) found from their data, the continuation ratio model provided the same conclusions as McCullagh's proportional odds model. The fit of the proportional odds model was to be slightly better than that of the continuation ratio model. According to Manor et al. (2000) results on the continuation ratio (single beta estimate) model were similar to those from the logistic regression models. However, again, the fit of the continuation ratio model was not as good as that for the proportional odds model. Harrell et al. (1998) also found a similar conclusion, in that the equal slope continuation ratio model fitted the data poorly compared to the constant slope proportional odds model. Armstrong and Sloan (1989), Scott et al. (1997), Bender and Benner (2000) and Cox and Chuang (1984) describe in some detail the similarities of the proportional odds and the continuation ratio models. Armstrong and Sloan (1989) illustrated how for their data, the standard error of the regression parameters from the continuation ratio were similar to the proportional odds, but were smaller than the cut-point specific binary logistic regressions.

Although the proportional odds and the continuation ratio models are similar, comparisons should be made with caution. With the continuation ratio model cut-point specific estimates can be considered independent whilst those from the proportional odds model cannot. The continuation ratios are based on conditional probabilities, whilst those from the proportions odds are based on cumulative ones. Ananth et al. (1997) states that the choice of the model whether the continuation-ratio model or and the proportional odds model, should be based on the goals of the statistical analysis.

Different slopes models

The polytomous model has often been cited in the literature as an ordinal regression model (Ananth et al., 1997; Lu, 1999), when in fact it is a model that fits multinomial categories, and does not account for ordinality of the y-response. It has quite often been cited as a model that forms a basis for the stereotype model (Hendrickx, 2000; Greenland, 1994; Anderson, 1984). In the study by Ananth et al. (1997) and by Lipsitz (1996), results produced using the polytomous model were similar to those provided by the adjacent-category (different slopes) model, although the underlying model assumptions were not the same.

Constant and different slopes models

The proportional odds model can be viewed as a model 'nested' within the unconstrained partial proportional odds model. Bender and Benner (2000) failed to compute the partial proportional odds models and used the separate binary logistic analysis instead. These authors argue that the separate models based on the binary outcome produce results that are very close to the unconstrained partial proportional odds model, and that careful application of these models can represent a simple and adequate tool to analyse ordinal data with non-proportional odds. Bender and Grouven (1998) found similar conclusions when comparing the polytomous regression analysis with the separate binary logistic regressions and the partial proportional odds models. In their paper, they concluded that the partial proportional odds models imply more efficiency, compared to the other models. There is a need therefore to find a way to fit the partial proportional odds model, especially since Ananth et al. (1997) in comparing the likelihood ratios between the two partial proportional odds and proportional odds models, found that the unconstrained model performed better than the constrained one and the proportional odds model.

Ananth et al. (1997) suggests that the formulation of the different slopes adjacent category model was more flexible when compared to the proportional odds or the continuation ratio (constant slope) models, in that the regression coefficient corresponding to a covariate is allowed to vary by every level of the response categories.

Cox and Chuang (1984) in comparing the polytomous model with the other models (proportional odds and the continuation ratio (constant slope) models) mention that with well-defined goals, one can choose a 'best' model to describe the data. In this study, the analyses from the three models provided similar results, and they complemented each other and in each case some extra information was provided about the data.

Manor et al. (2000) found that with respect to the goodness-of-fit the polytomous model was generally a better fit than the constant slope adjacent category, constant slope continuation-ratio and the proportional odds models.

Greenwood and Farewell (1988) illustrated how a one-dimensional and possibly a two-dimensional stereotype model could be considered given their data. In summary, the two-dimensional stereotype model was rated better than the proportional odds, continuation ratio and the one-dimensional stereotype model as the effect of the response was found to differ between the response categories and covariates. However, it has to be borne in mind that their deduction of a two-dimensional stereotype model being a 'better' model was purely based on the observation of the parameter estimates and odds ratios. Holtbrügge and Schmacher (1991) carried out a simulation study and found that the results indicated less bias in the estimates produced from the proportional odds model compared to the stereotype model. An improvement occurred in the estimates using the latter model, if adjacent categories were amalgamated. An explanation for the increase in bias for the parameters of the stereotype model was largely attributed to the non-linear relationship and correlation between the regression coefficient β_k (k=1,...., p) and ϕ_j (j=1,..., c-1). This study was the first to highlight the conditional relationship of the estimated parameters, and this was later picked-up by Greenland (1994) and again by Hendrickx (2000). Greenland (1994) gave a sound explanation of the use of the stereotype model, and showed its superiority in the context of an ordinal scale where the categories were not assumed to have an ordinal structure. This study illustrated that the stereotype model was the only model among the proportional odds, continuation ratio and the stereotype model itself, to reproduce data patterns that may be important on a priori grounds. In this paper, however, only one example was discussed and this example consisted of one covariate with a three-point ordinal response.

The stereotype model compared to other models, is less parsimonious than the proportional odds model, since it has extra c-2 parameters for the scaling metric.

The stereotype model can be considered as a constrained multinomial model. In fact, if there is only one predictor, the stereotype model is simply a re-parameterisation of the polytomous model: the goodness-of-fit, predicted values and so on are all identical. However, this is not true in the case of a stereotype model and a polytomous model with two or more covariates. As the covariates increase in the models, the number of parameters estimated in the stereotype model does not increase as rapidly as in the polytomous model.

Joffe and Greenland (1995) states that the adjacent-category model is a special case of the stereotype model in which the scores {ϕ_js} are fixed a priori.

3.12 Interpretation of Ordinal Regression Models

The constant slope models (i.e. proportional odds, constant slope adjacent category and constant slope continuation ratio models) are, by far, the easiest to interpret the ordinal scale, as there is only one single measure (odds ratio) used to summarise the entire response categories. However, the models assumptions are often difficult to satisfy (Ananth et al., 1997; Bender et al., 1997). At the other extreme, are the different slope models (polytomous, different slope adjacent category, different slope continuation ratio and the stereotype models) - for which there are as many odds ratios used to interpret the response scale, as there are cut-points. Although the goodness-of-fit of these latter models is adequate compared to the constant slope models (Ananth et al., 1997), the interpretation of the odds ratios is however more difficult due to the large number of odds ratios. The partial proportional odds and the partially constrained continuation ratio models are halfway between these two extreme cases. If there is evidence of some homogeneity in the cut-point specific regression parameters from the different slope models, then these models allow the flexibility of fitting the same odds ratios over the latter cut-points (Peterson and Harrell, 1990; Cole and Ananth, 2001), resulting in some easy of interpretation of the response scale. The added advantage of these models is that the goodness-of-fit can also be retained, whilst looking to constrain regression parameters that are homogenous over the cut-points.

3.13 Summary

The following models were reviewed in this chapter:-

- Polytomous Model;
- Proportional Odds Model;
- Unconstrained Partial Proportional Odds Model;
- Constrained Partial Proportional Odds Model;
- Adjacent Category Model;
- Continuation Ratio Model;
- Stereotype Model.

In relation to the aims of this thesis, the relevant statistical aspect of each model were summarised in Table 3.2.

In addition to the latter the following are also highlighted:

(a) Model Fitting

- The stereotype model when fitted using Hendrickx (2000) macros becomes a non-linear regression model.

- No statistical software exists to fit the partial proportional odds models.

- The unconstrained continuation ratio model can be fitted using the method described by Cole and Ananth (2001). However, this method involves manipulation of the data, prior to analysis.

(b) Model checking

(i) *Model Assumptions:* The assumption of proportional odds can be checked using the likelihood ratio test/ χ^2-score test/ smooth residual plots/ Wald test/ Brant's Method. The assumption of parallel slopes for the continuation ratio model can be checked using the χ^2-score test. However no test is given to check the assumption of a constant slope for the adjacent category model.

(ii) *Goodness-of-fit:* The methods used to assess the goodness-of-fit of an ordinal regression model are similar to those of the binary logistic regression model. The observed and predicted observations can be compared using the Pearson's chi-squared/likelihood ratio test. Given cell counts are less than 5 in frequency, the generalisation of the Hosmer-Lemeshow goodness-of-fit statistic for ordinal data as derived by Lipsitz (1996) can be applied.

Although the stereotype model can be fitted using specially devised macros, the goodness-of-fit cannot be checked easily. In fact, both Henrickx (2000) and Lunt (2001) make no mention regarding the goodness-of-fit of the stereotype model. Also, the comparison of different forms of the stereotype model (i.e. main effects and saturated models) is difficult, as it cannot be shown that the log likelihoods of these models follow a χ^2 distribution asymptotically.

(iii) *Residual Analysis:* For ordinal regression models methods used to assess the individual residuals and observations are very limited and under-developed. Individual binary logistic regression models based on the cut-points can be fitted for the polytomous models and the residual analysis can be carried out using the latter models and results integrated to give an overall conclusion. For other models simple method such as visually comparison of the observed and expected cell frequencies/probabilities are used to examine individual extreme values or 'outliers'. The residuals from the logits can also be used.

(iv) *Comparison of the model fits:* The AIC/SC statistics are applied to compare the different ordinal regression models. These statistics are computed using the log-likelihood values of the models. In addition to these the partial proportional odds models are compared to one another and to the proportional odds model by the use of the score test and assessing the log-likelihoods.

(c) Several Covariates

The asymptotic relative precision (ARP) can be computed to assess the effect of the standard error given a single and adjusted covariate using binary logistic regression models. The ARP is an indication of how precise the estimate is when fitted singularly compared to when fitted with other covariates. The ARP of a covariate has not been assessed in the context of ordinal regression models.

(d) Interaction terms

The interaction terms in the presence of several ordinal response categories has not been researched very well in the literature. The interpretation of an interaction term in an ordinal regression model would be different to that presented in a binary logistic regression model due to the presence of several *y*-response categories and also for some models there are constant regression parameters whereas for others there are different parameters over the cut-points.

The use of interaction terms in a stereotype model has not been explored in the literature. The model using Hendrick's (2000) macros is fitted as non-linear and therefore one would need to be cautious when comparing main effects model and interaction model, as it cannot be shown that the log-likelihoods of these models follow a χ^2 distribution asymptotically.

(e) Sparse Data

Small-sample test are still under-developed and computationally not feasible.

(f) When to use a given ordinal regression model

In the literature the proportional odds model is quite frequently used in the context of grouped continuous ordinal data, and the polytomous and stereotype models are often referred when a truly discrete ordinal scale is present. However, given these types of data, other models are also applicable and there are no clear guidelines regarding the use of certain ordinal regression models given certain data.

Table 3.2: Summary of Ordinal Regression Models

Model	Prime Feature	Model form	Link Function	Odds Ratios	Software (SAS)
Polytomous	Assessing the response in a category versus that in a referent category. This model is does not account for ordinality of the *y*-response	Different slope parameters model	$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_c)}\right]$	Several cut-point specific odds ratios	PROC CATMOD using the *logit link*
Proportional Odds	Used primarily in the presence of an underlying continuum	One single parameter model	$\ln\left[\frac{\Pr(Y_i \le y_j)}{\Pr(Y_i > y_j)}\right]$	Single odds ratio	Use of PROC LOGISTIC
Unconstrained Partial Proportional odds	Used primarily in the presence of an underlying continuum	Single (where proportional odds is satisfied)/several (where proportional odds is not satisfied) slopes parameter model	$\ln\left[\frac{\Pr(Y_i \le y_j)}{\Pr(Y_i > y_j)}\right]$	Single/several cut-point specific odds ratios	No longer possible to compute in SAS
Constrained Partial Proportional odds	Used primarily in the presence of an underlying continuum	Single (where proportional odds is satisfied)/several with constraints (where proportional odds is not satisfied) slopes parameter model	$\ln\left[\frac{\Pr(Y_i \le y_j)}{\Pr(Y_i > y_j)}\right]$	Single/several cut-point specific (allowing for a trend) odds ratios	No longer possible to compute in SAS
Adjacent Category	Used when one is interested in assessing the outcome of two neighbouring categories	(i) Single slope parameter model; (ii) Different slope parameter model	$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_{j+1})}\right]$	(i) Single odds ratio; (ii) Several cut-point specific odds ratios	(i) Use of PROC CATMOD for constant slope model; (ii) Not known for different slopes model
Continuation Ratio	Particularly relevant when assessing outcomes which are irreversible	(i) Single slope parameter model; (ii) Different slope parameter model	$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i > y_j)}\right]$	(i) Single odds ratio; (ii) Several cut-point specific odds ratios	(i) PROC LOGISTIC using the clog-log to fit the constant slope; (ii) PROC LOGISTIC- separate binary logistic models to fit the separate slopes
Stereotype	Primarily develop to model truly discrete outcomes	Combination of single (slope) and several (ordinality) parameter model	$\ln\left[\frac{\Pr(Y_i = y_j)}{\Pr(Y_i = y_c)}\right]$	Several cut-point specific odds ratios	Hendrickx's (2000) and Lunt's (2001) macros

CHAPTER 4 - STUDY DESIGN AND DATA

4.1 Aims of this Chapter

The main objectives of this chapter are: -

(i) to describe the aims and design of the Medical Research Council Cognitive and Function Ageing Study (MRC CFAS);

(ii) to identify the two health related quality of life measures, one with an underlying continuum (group continuous) and one truly discrete (assessed) from the data;

(iii) to identify a set of covariates for each measure, for the purpose of fitting the regression models.

Section 4.2 details the aims of the study and section 4.3 describes the study design. The number of subjects eligible for the study is given in section 4.4. The data that were collected in the study in relation to the two different types of health related quality of life instruments are outlined in section 4.5. In this section detail is also given of how the covariates were determined for fitting the regression models.

4.2 Aims of the study

(a) The study

The data used to implement the regression models were obtained from a multi-centre study of cognitive function and ageing. This study known as the Medical Research Council Cognitive and Function Ageing Study (MRC CFAS) commenced as a longitudinal epidemiological study. It was funded by the Medical Research Council and the Department of Health.

(b) Its aims

The main aim of the study was to examine the natural history of cognitive decline and dementia in the elderly population and to evaluate the degree of disability associated with any decline.

4.3 Study design and the data collected

(a) Study Design

Six centres in United Kingdom were chosen to study the variation throughout the country in the prevalence and incidence of dementia and cognitive decline. These included two rural areas – Cambridgeshire and Gwynedd and four urban areas – Liverpool, Newcastle, Nottingham and Oxford. These areas captured the main natural variation in urban-rural differences, the North-South and East-West gradient and variations in socio-economic levels and in rates for other chronic diseases, such as stroke, ischaemic heart disease and cancers of the lung, stomach and breast (Swerdlow and dos Santos Silva, 1993). The Liverpool component, however, was funded and started before planning of the MRC CFAS study was completed, and as a result its design differed in some detail, compared to that of other centres. For this reason, this centre was excluded from the analysis and the five remaining centres were used.

The cognitive function of a population of individuals aged 65 years and over was examined on two occasions, the first at the *prevalence visits*, for prevalence estimates and the second at the *incidence visits* for incidence estimates. On both occasions there was a *screening interview*, which was followed, some weeks later, by an *assessment interview*. Subsequently other visits followed (annual follow-up visits and the combined screen and assessment interviews), but these have not been detailed here, as the analysis in this thesis is based on data collected at the initial prevalence screening interview.

(b) Data collected

The MRC CFAS prevalence screen interview was designed to define the population studied. The full interview consisted of approximately 207 questions that related to orientation, socio-demography (including age, sex, marital status, education and social class), social contacts, general health, memory, sleeping problems, smoking, drinking and medication. The data on the degree of social integration and contacts were also assessed. Questions on the Activities of Daily Living were asked in order to establish the level of disability of a respondent. Cognitive function was measured using the Mini-Mental State Examination (Folstein et al., 1975) and data were collected to ascertain core risk factors as measured in the EURODEM studies of Alzheimer's disease, vascular dementia and cognitive decline (Van Duiju and Hofman, 1992).

4.4 Number of subjects sampled

At each centre random samples were selected of sufficient size to yield 2500 interviews from individuals aged 65 years and over, with equal numbers in the age groups 65-74 years old and 75 years old and above. The total sample available at baseline was 20 234 for the five centres and there were 17 608 respondents identified as being eligible. Of these 13 006 were interviewed at the initial visit and were regarded as the "achieved sample" (MRC CFAS[2], 1998). This sample provided the basis for the analysis.

4.5 Health-related quality of life measures and the covariates

Several versions of the data were collected and version 4.1 was used for the analysis in this thesis.

The aim here was to isolate outcomes which examined an aspect of quality of life, or quality of life itself, and which were measured on an ordinal categorical scale.

A search was made to find two measures of health related quality of life, one which was of an 'assessed' type (i.e. with truly discrete categories) and one which was of a 'group continuous' type (i.e. where an underlying continuum variable was present). Appendix I details a list (Listing 1) of the ordinal outcomes that were recorded at the prevalence screening and the choice of the two measures was made from this list.

4.5.1 Assessed ordinal outcome –Health Status measure

From Listing 1a (Appendix I), the most appropriate 'assessed' outcome in the context of health-related quality of life was the Health Status question (which has been adapted from the health status question from the SF-36 Health Survey (Ware et al., 1998)): "Would you say that for someone of your age, your own health in general is: '*Excellent*' (0), '*Good*' (1), '*Fair*' (2), '*Poor*' (3), '*No answer*', '*Not asked*' and '*missing*'?"

4.5.1.1 Choice of the covariates for the Health Status measure

(a) In the literature

Much has been published using the MRC CFAS data, but there are no publications that use the Health Status question as a primary or a secondary end-point. Thus, the choice of covariates for fitting the regression models had to be based on what was relevant in the literature. Table 2.1, in the Background section lists some studies that are cited in the literature where health status was used as an outcome response. From this table, it is evident that the most common independent risk factors that related to health status include age, sex, education, marital status, smoking and alcohol consumption. Assessments that are less common but still relevant include exercise, the BMI index, chronic diseases (such as heart disease/ heart attacks, diabetes etc.) and depression.

(b) Criteria for the chosen covariates

The MRC CFAS study did not record all these variables; for instance, exercise, BMI index and depression were not directly assessed. The variables that were examined were age, sex, level of education, marital status, smoking, alcohol consumption, heart attacks, angina and diabetes. The intention was not to use all these variables in the fitting of the regression models. The decision of the choice of the covariates was driven by the need to assess the effect of two or more covariates, and in particularly to examine the covariate parameter estimates with respect to one another. As stated in section 3.7, Robinson and Jewel (1991) have reported that given two covariates x_{i1} and x_{i2}, and the response Y_i, then if conditions (i) and (ii) are satisfied then the variance of x_{i1} will always be smaller than the variance of x_{i1} having adjusted for x_{i2}. This has not been assessed using ordinal regression models and therefore the choice of the covariates was based on satisfying either criteria (i) or (ii).

4.5.1.2 Exploratory analysis using the Health Status measure and the covariates

An exploratory analysis of the Health Status outcome with respect to the covariates was carried out in order to identify the covariates that satisfied one of the above criteria.

There were 309 respondents who had missing data on their health status, 61 respondents did not know the rating of their health status and 14 subjects were not asked the Health Status question. Thus, in total there were 384 observations that were disregarded on the health status assessment, as these categories could not be incorporated into the ordinal analysis. A total of

12622 observations were used for the analysis. Centre effect, although not considered in the above list was also assessed in relation to health status.

4.5.1.3 Results of the exploratory analysis using the Health Status measure and the covariates

Initially each of the covariates was cross-tabulated with the one another and the response and the association examined. The covariates used are listed in Appendix I in Table 1a.

(a) Association of the covariates with one another

Table Ib (in Appendix I) illustrates the association of each of the covariates with one another (including the health status outcome) using p-values as obtained from the Pearson's chi-squared test statistic. Missing response data or a response of '*no answer*' or '*not asked*' were disregarded and only complete observations were used for the covariates to assess the association. Although the Pearson's chi-squared test statistic does not account for the ordinal categories, it does give a crude indication of how well the variables are related to one other. From Table Ib (in Appendix I) there is indication many of the covariates were strongly associated with one another satisfy criteria (i) above. The choice of the covariates was difficult using this criterion.

(b) Association of the response with the covariates

The association of two given covariates with respect to the response was assessed using the Cochran-Mantel-Haenzel (row mean score) statistic as presented by Mantel (1963) for assessing criteria (ii). This statistic examines the association of the response and one given covariate, whilst adjusting for the effect of the other covariate by treating it as a stratification variable. The ordering of the response variable is taken into account by assigning scores to the response categories, forming means and then examining location shifts of the means across the sub-populations. A significant association between general health status of the respondent and whether he/she has had a heart attack (after controlling for whether he/she smokes or not – Q_{SMH}=190.767 on 1 d.f.; p =0.001) was provided, when examining the association of two covariates with respect to the response. Likewise there was evidence of a notable association between the health status and whether or not a respondent smoked (after having accounted for the fact that a respondent may or may not have had a heart attack -Q_{SMH}=4.212 on 1 d.f.; p =0.04). These results indicated that the adjusted heart attack covariate was strongly associated with health status, and the association with the smokers/non-smokers, although evident, was not as strong. These two covariates did not satisfy criteria (i), however they did satisfy the above criteria (ii) and were therefore examined further and consequently used for fitting the regression models.

(c) Health Status versus the 'smoke' and 'heart attack' covariate

The two covariates were individually cross-tabulated against the health status score. There was an indication that those who did not smoke were more likely to have '*excellent*' or '*good*' health status, where as smokers were more likely to have '*poor*' health (Q_{SMH}= 4.607 on 1 d.f; p =0.0318). Also, there was a strong association with health status and 'heart attack' (Q_{SMH}=126.33 on 1 d.f.; p <0.0001). Those who had not suffered from a heart attack were more likely to have '*excellent*' and '*good*' health, whilst those who had suffered from a heart attack were more likely to have '*fair*' or '*poor*' health.

Both the 'smoke' and 'heart attack' were cross-tabulated with the health status score and presented in Table 4.1. There were 12535 complete observations available on both covariates and the response.

4.5.2 Group continuous ordinal outcome- Townsend Disability Scale

Two variables were recorded as grouped continuous ordinal outcomes: the Mini-Mental State Examination (MMSE) (Folstein, 1985) and the 6-group Townsend Disability Scale (Townsend, 1979). The Mini-Mental State Examination is a well-validated scale but no corresponding validated ordinal categories exist for this scale. However, studies have been published where results are summarised using an ordinal form of the MMSE (score <21, 22-25, 26-30) (MRC CFAS[2], 1998).

The 6-group Townsend Disability Scale is a short index of activities that assess physical ability in social terms (see Listing Ib in Appendix I). The scale consists of nine questions or 'items', and each item is rated using a level of difficulty: 0-'yes, with no difficulty', 1-yes, with some difficulty' and 2-'no, needs help'. The scale gives equal weighting to each item and a summary of the level of disability is provided by the total score. The total disability score can be categorised into six groups: a total score of 0 is regarded as indicating no disability; 1-2 'slightly disabled', 3-6 'some disability', 7-10 'appreciable disability', 11-14 'severe disability' and 15-18 'very severe disability'. Although these ordinal categories have not been validated, they are published (Townsend, 1979). In the context of the MRC CFAS data, work has been carried out and published using the disability score, but not these ordinal categories (MRC CFAS[1], 1998). As the ordinal categories based on the Townsend disability score are more reliable than those of the MMSE, it was decided that the 6-group ordinal form of the Townsend Disability Scale be used in the fitting of the regression models.

4.5.2.1 Choice of the covariates for the Townsend Disability Scale

The choice of the covariates using the 6-group Townsend Disability Scale was based on the published work using the disability score (MRC CFAS[1], 1998; MRC CFAS, 2000).

(a) Adjusting covariates

In the MRC CFAS[1] (1998) publication, logistic regression models were fitted and it was found that the centre effect was small. Also each covariate was adjusted for 'age group' (under 70, 70-74, 75-79, 80-84, 85+ years) and 'sex'. In the MRC CFAS (2000) publication, summary statistics and exploratory plots of the estimated years of disability and prevalence of disability were provided for each sex, age group (65-74, 75-84, 84 + years) and social class (I, II, III, IV, V). From these two publications, the most common variables, namely 'age group' (under 70, 70-74, 75-79, 80-84, 85+ years) and 'sex' (male/female) were taken as the adjusting covariates when fitting the regression models. Note that the choice of the five age-group categories for the analysis in this thesis was analogous to that in the MRC CFAS[1] (1998) publication. The reason for this choice was so as that results from the models computed in this thesis could be compared with those in the latter publication.

(b) Causal covariates

The causal covariates which were listed in the MRC CFAS[1] study (1998), included centre; marital status (single/with partner, widowed, divorced/separated); social class (I and II, III (non-manual), IV and V, armed forces); full-time education (<9 years, 9 years, 10-12, ≥ 13 years); type of accommodation (house, flat or granny flat, warden-controlled flat, home or hospital); deterioration in eyesight (none, some, marked/ blind); and deterioration in speech (none, some, marked/dumb).

The choice of a third covariate came from this group of variables, and it was driven by the fact that it had to be binary. The reason for this was that very early on in the analysis, it was realised that the fitting of ordinal regression models becomes quite complicated as more and more covariates are added into a model. To allow a covariate to be added into a model adequately having adjusted for 'age group' and 'sex', there was a need to have this covariate as simple as possible. As a result 'level of education (in years)' was chosen. For the analysis in the MRC CFAS[1] (1998) paper, the latter variable indicated that the odds of disability occurring for those with (< 9 years) education was 1.03 (95% C.I.:0.9, 1.1) compared to those with 9 years education; the odds of disability occurring for those with 10-12 years education was 0.92 (95% C.I.: 0.9, 1.0) compared to those with 9 years education and the odds of disability occurring for those with 13 or more years of full-time education was 0.76 (95% C.I: 0.7, 0.8) compared to those with 9 years of full-time education. From these results the 95% confidence intervals indicated that those with <9 years of full-time education were not statistically different than those with 9 years education (since it contains '1'). Similarly those with 10-12 years of full-time education did not statistically differ in terms of their odds compared to those with 9 years of full-time education. This provided a justification to collapse the three full-time education groups (<9, 9, 10-12 years) into one group. A new full-time education variable was formed which was based on two categories (less than 13 years of full-time education and 13 or more years of full-time education).

4.5.2.2 Exploratory analysis using the Townsend disability score and the covariates

Initially the disability score (i.e. continuous score) was tabulated against 'sex' and 'age groups'. The summary statistics of this continuous form of the response were compared against the summary statistics of the 6- group ordinal categories of the Townsend Disability Scale. This clarified that no major differences had occurred in the response data when they had been transformed from the continuous rating score to the ordered categories. This was verified as there is some indication in the literature that often continuous variables when converted into ordered categorical variables by grouping values, can introduce an extreme form of measurement error with an associated loss of power (Agresti, 1999).

For the 6-grouped Townsend disability ordinal score, 572 respondents did not have an assessment. This was as a result of missing items or items on the scale that were not answered or asked. In total 12434 subjects provided a Townsend disability score. The covariates that were assessed in relation to the 6-grouped ordinal categories included 'centre', 'sex', 'age groups' and 'level of education'. There were no missing data for these variables.

(a) Townsend disability score by centre

The number and percentage of respondents for each centre and each of the six ordinal response categories were tabulated. The distributions across the ordinal disability categories for the centres were very similar. Each distribution was right-skewed, with larger proportion of subjects with no disability. Although no statistical testing was carried out, it was decided, as in the MRC CFAS[1] (1998) study that there was little centre effect. For this reason this variable was not considered further.

(b) The 6-grouped Townsend Disability Scale collapsed into a 5-grouped Townsend Disability Scale

The 6-grouped ordinal Townsend disability categories were cross-tabulated for 'sex,' 'age group' and 'level of education'. The frequency table indicates that there were some missing data, which had been generated as a result of the cross-tabulation. In particular there were no respondents for the categories: (a) males, < 70 years old with 13 or more years of full-time education for very severe response; (b) females, < 70 years old with 13 or more years of full-time education for severe response, and (c) males, 80-84 years old, with 13 or more years of full-time education for very severe response. The missing data would lead to computation difficulties when fitting the models. If replaced by zero there would still be some problems, as the denominator of the log odds would have to take on a zero marginal probability. For this reason, the last two categories ('*severe*' and '*very severe*') of the response were grouped into one. As of this point onwards, only the 5-grouped ordinal form of the Townsend Disability Scale, i.e. '*none*', '*slight*', '*some*', '*appreciable*', '*severe + very severe*' was considered. The combining of the categories was also done in the MRC CFAS[1] (1998) publication.

(c) The 5-grouped ordinal Townsend disability categories tabulated against each single covariate (age groups, sex and level of education)

Each main covariate was tabulated against the five ordinal disability groupings. There was significant statistical association between the covariates and level of disability ('sex' v. Townsend disability categories: Q_{SMH}=577.45 on 1 d.f.; p <0.0001; 'age group' v. Townsend disability categories: Q_{SMH}=2214.18 on 1 d.f.; p <0.0001; 'full-time education' v. Townsend disability categories: Q_{SMH}=51.98 on 1 d.f.; p <0.0001). This indicated that each covariate was strongly related to Townsend disability score.

(d) The 5-grouped ordinal Townsend disability categories tabulated against two covariates (sex and level of education, age and level of education)

Two covariates were tabulated against the Townsend disability scores. It was not considered relevant to tabulate the adjusting variables 'sex' and 'age group' with the response variable, as focus was centred round 'full-time education'. It was evident that when taking each 'age group' as a stratum, there was significant association between 'full-time education' and the levels of disability (Q_{SMH}=66.90 on 1 d.f.; p <0.0001). Also, within each 'sex' category there was an indication of significant association between the level of 'full-time education' and the level of disability (Q_{SMH}=51.98 on 1 d.f.; p <0.0001).

(e) The 5-grouped ordinal Townsend disability categories tabulated against the three covariates (sex and age groups and level of education)

Table 4.2 provides the cross-tabulation of 'sex' given each 'age group', each level of 'full-time education', for the 5-ordinal response categories. This table summaries the number and proportion of respondents within each level of the covariate, allowing for the other covariates. From this table, there are some very low frequency counts and proportions.

4.6 Summary

(i) Two quality of life measures were chosen: -

- Health Status – a discrete ordinal categorical scale- ('*Excellent*' (0), '*Good*' (1), '*Fair*' (2), '*Poor*' (3));

- Townsend Disability Scale – a scale with an underlying continuum- ('*none*' (0), '*slight*' (1), '*some*' (2), '*appreciable*' (3), '*severe + very severe*' (4)).

(ii) For the purpose of fitting the regression models the covariates for each health related quality of life measure were: -

Quality of life measure	**Covariates**
Health Status question	Heart attack (yes/no);
	Smoke (yes/no)
Townsend Disability Scale	Age group (< 70, 70-74, 75-79, 80-84, 85+ years);
	Sex (male/female);
	Full-time education (<13 or ≥ 13 years)

(iii) The Health Status score has not been analysed in any previous publication. The disability scale has been previously analysed in the MRC CFAS[1] (1998) publication. However, the ordinal form of the scale has not been assessed.

(iv) The data on the health status did not seem to pose any problems – the frequency cells were large with no missing data and the distribution of the response with respect to the covariates was well spread out. For the 5 grouped Townsend disability score categories, the data were skewed with respect to the covariates and there were some cell frequencies with very little data.

Table 4.1: A frequency table displaying the number and percentage (%) of respondents within each Health Status score, smoking and heart attack categories

Have you had a heart attack?	**Do you smoke?**	**Health Status categories**				
		Excellent	***Good***	***Fair***	***Poor***	***Total***
Yes	***Yes***	27 (11.1%)	76 (31.3%)	101 (41.6%)	39 (16.1%)	243
Yes	***No***	83 (7.9%)	406 (38.9%)	442 (42.3%)	114 (10.9%)	1045
No	***Yes***	402 (19.0%)	1050 (49.6%)	522 (24.6%)	145 (6.8%)	2119
No	***No***	1959 (21.5%)	4521 (49.5%)	2243 (24.6%)	405 (4.4%)	9128

Note: that the parentheses reference the percentages based on the row marginal totals

Table 4.2: A frequency table illustrating the number and percentage of respondents within the Townsend Disability Scale categories for age group, sex and full-time education

Age group (Years)	Sex	Level of education (Years)	5-grouped Townsend Disability Scale categories					
			None	*Slight*	*Some*	*Appreciable*	*Severe+ very severe*	*Total*
<70	*Male*	*< 13*	803 (62.6%)	252 (19.6%)	147 (11.5%)	41 (3.2%)	40 (3.1%)	1283
		≥13	96 (68.1%)	22 (15.6%)	16 (11.4%)	5 (3.6%)	2 (1.4%)	141
	Females	*< 13*	731 (47.1%)	385 (24.8%)	252 (16.2%)	104 (6.7%)	80 (5.2%)	1552
		≥13	107 (61.1%)	42 (24.0%)	17 (9.7%)	8 (4.6%)	1 (0.6%)	175
70-74	*Male*	*< 13*	677 (55.7%)	279 (22.9%)	149 (12.3%)	59 (4.9%)	52 (4.3%)	1216
		≥13	92 (63.5%)	31 (21.4%)	14 (9.7%)	4 (2.8%)	4 (2.8%)	145
	Females	*< 13*	580 (36.9%)	415 (26.4%)	339 (21.5%)	155 (9.9%)	85 (5.4%)	1574
		≥13	73 (43.5%)	50 (29.8%)	30 (17.9%)	9 (5.4%)	6 (3.6%)	168
75-79	*Male*	*< 13*	399 (39.9%)	255 (25.5%)	193 (19.3%)	68 (6.8%)	86 (8.6%)	1001
		≥13	56 (45.2%)	35 (28.2%)	22 (17.7%)	6 (4.8%)	5 (4.0%)	124
	Females	*< 13*	357 (23.5%)	376 (24.7%)	416 (27.4%)	193 (12.7%)	178 (11.7%)	1520
		≥13	54 (32.5%)	43 (25.9%)	45 (27.1%)	14 (8.4%)	10 (6.0%)	166
80-84	*Male*	*< 13*	175 (26.8%)	145 (22.2%)	170 (26.1%)	83 (12.7%)	79 (12.1%)	652
		≥13	28 (29.5%)	26 (27.4%)	30 (31.6%)	5 (5.3%)	6 (6.3%)	95
	Females	*< 13*	159 (12.8%)	206 (16.6%)	338 (27.2%)	244 (19.7%)	294 (23.7%)	1241
		≥13	18 (15.1%)	29 (24.4%)	37 (31.1%)	14 (11.8%)	21 (17.7%)	119
≥85	*Male*	*< 13*	34 (10.9%)	50 (16.1%)	66 (21.3%)	73 (23.6%)	87 (28.1%)	310
		≥13	8 (24.2%)	6 (18.2%)	11 (33.3%)	5 (15.2%)	3 (9.1%)	33
	Females	*< 13*	27 (3.2%)	76 (9.0%)	174 (20.7%)	196 (23.3%)	368 (43.8%)	841
		≥13	5 (6.4%)	9 (11.5%)	21 (26.9%)	19 (24.4%)	24 (30.8%)	78

Note: that the parentheses reference the percentages based on the row marginal totals

CHAPTER 5 - STATISTICAL ANALYSIS

5.1 Aims of this Chapter

The main objective of this chapter is to illustrate the statistical methodology and model fitting used for the regression models, using the two quality of life measures (Health Status and Townsend Disability Scale) identified in Chapter 4. The following chapters - Chapter 5 and Chapter 6 -provide the methods for goodness-of-fit and the results obtained from the model fitting respectively. Inevitably, there is some overlap in these three chapters. However, due to the large number of models fitted, attempt has been made to separate out the sections of methodology, model checking and results as much as possible, so as that it is evident how each model was fitted and checked for goodness-of-fit and what the final results were in terms of the odds ratios.

(i) Regression models fitted

In this chapter, detail is focused on the ordinal regression models. The models fitted were:-

- Linear Regression Models;
- Binary Logistic Regression Models;
- Polytomous Models;
- Proportional Odds Models;
- Unconstrained Partial Proportional Odds Model;
- Constrained Partial Proportional Odds Model;
- Adjacent Category Models (constant slope and different slopes);
- Continuation Ratio Models (different slopes and fully constrained);
- Stereotype Models (linear and non-linear forms).

(ii) Issues addressed

There are two main aspects addressed in this chapter.

(a) Various issues related to fitting ordinal regression models have received little attention in the literature and are relevant to the data in this thesis (as identified in Chapter 3). In this chapter, these issues are considered in some depth. These include:

i. fitting several variables in ordinal regression models;
ii. fitting interaction terms and assessing the significance of these in the model;
iii. fitting the partial proportional odds models (as no standard method exists to fit these models);
iv. fitting the unconstrained continuation ratio model. This can be fitted using the method described by Cole and Ananth (2001). However, the method used in this chapter is more computationally efficient.
v. The stereotype model as fitted using Hendrickx's (2000) and Lunt's (2001) macros is a non-linear model. In order to check the goodness-of-fit of this model, one can take the ordering constraint parameters (ϕ_j) obtained from the macros and incorporate them into the stereotype model (3.23) estimating what we shall call a *linear stereotype model*. This latter model allows one to assess the goodness-of-fit of the non-linear form.
vi. Model comparison of the different forms of the stereotype model can be carried out using the *bootstrap technique*. This method has not been applied any way in the literature to compare models, and is similar to the Monte Carlo simulation used by Greenland (1994).

(b) The disability score (continuous form of the Townsend disability score) has already been analysed in the MRC CFAS[1] (1998) publication. In the latter, the disability score for each subject was taken as a proportion of the total score and the binary logistic model was used to fit the data. The interpretation of the results was not clear, so the analysis was replicated here using 'sex', 'age group' and 'full-time education' as covariates. Also the results from this analysis were compared to the results from the linear regression and ordinal regression models.

(iii) Computing the regression models

The statistical software package *SAS* (version 8.01) and *Stata* (version 7.0) were used to compute the above models.

(a) Linear Regression Model: *PROC GLM* was used to fit the linear regression models for both datasets using standard methods (see Appendix II).

(b) Binary Logistic and Ordinal Regression Models: *PROC CATMOD* was used to compute the binary and ordinal regression models for both datasets. However, the standard procedure in *SAS* was not applied. Instead a method cited by Stoke et al. (1995) for fitting models was considered. These investigators pursued to fit the main effects proportional odds and polytomous models using a method based on first principles and this method can be adapted for all models based on ordinal categories. In this chapter this method is considered in depth in relation to the ordinal regression analysis.

Consider the general form of an ordinal regression model, given in equation 3.5. Note that in this equation the observed values of the dependent variable are not in the equation. They are linked to the model by the multinomial distribution. Thus in cell i if we observe n_{ij} subjects out of n_{i+} then we assume that the n_{ij} are distributed multinomially with probability π_{ij}. The parameters in the model are estimated by either maximum likelihood or weighted least squares. Of course we do not know the population values π_{ij} and in the modelling process we substitute into the model the estimated or fitted values. The observed sample proportions are referenced as p_{ij}, and replace the π_{ij}. The $\hat{\pi}_{ij}$ vector is an estimate of π_{ij} and estimates the probability of an event from the model. These are the predicted or fitted values for n_{ij}. A good model will give predictions $\hat{\pi}_{ij}$ close to the observed proportions n_{ij}/n_{i+}. Thus equation 3.5 models the vector of logits $F(\hat{\pi}_i)$, such that $F(\hat{\pi}_i) = \{f_{i1}(\hat{\pi}), f_{i2}(\hat{\pi}), \ldots. f_{i(c-1)}(\hat{\pi})\}$ where i=1... r. We can re-write this complete vector of logits as $F(\hat{\pi}) = (F(\hat{\pi}_1), F(\hat{\pi}_2) \ldots. F(\hat{\pi}_r))$. We refer to the predicted logits collectively using the latter notation and there are, at the most, $r(c$-1) of these logits or *response functions*. The observed sample response functions are $F(p)$. Then, in brief, the method involves computing the following steps (details are given in Appendix II: section 2):

i. the observed sample marginal probabilities (p_{ij}) and their variance covariance matrix are obtained.
ii. The observed sample response functions $F(p)$ are computed together with their variance covariance matrix.
iii. The parameter estimates of an appropriate ordinal regression model are obtained using weighted least squares or maximum likelihood estimation, given a design matrix and the observed marginal probabilities and response functions obtained in (i) and (ii).

iv. Once the parameter estimates are obtained $F(\hat{\pi}) = X\beta$ and the variance of $F(\hat{\pi}) = XV(\beta)^{-1}X$ are computed, where β denotes the vector of the estimated parameters and X is the model specification matrix or design matrix (containing 0, -1 and 1).

In this thesis, using this method, the regression models outlined above are fitted to the Health Status data (as in section 5.2) and to the Townsend Disability Scale data (as in section 5.3).

5.2. Health Status data

The 'smoke' and 'heart attack' covariates were taken as categorical for all the regression models fitted. These covariates were chosen using the criteria specified in 4.5.1.1 (b). Therefore the results from the models (with the exception of the linear regression model) were used to obtain the Asymptotic Relative Precision (ARP) statistics, which gave an indication of the precision of the estimate of the regression parameter when 'smoke' was fitted as a single covariate, compared to when it was adjusted for using 'heart attack'. The binary and ordinal regression models were therefore fitted with (a) unadjusted 'smoke' covariate; (b) adjusting 'smoke' for 'heart attack'. Below only the analysis for (b) is illustrated, as this is the more complicated of the two.

Note that the response categories '*Excellent* (0)', '*Good* (1)', '*Fair* (2)' and '*Poor* (3)' were recoded as '*Excellent* (1)', '*Good* (2)', '*Fair* (3)' and '*Poor* (4)' due to the fact that *STATA* becomes 'disabled' when fitting models where the response categories are coded as '0'.

5.2.1 Linear Regression Model

(a) Model assumptions

Prior to fitting linear regression models, it is usual to check the model assumptions. These assumptions are as follows:

(a) a Normal distribution is assumed for the outcome given the covariates;
(b) the y-observations are assumed to have equal variances and to be independent;
(c) the relationship of the response and each covariate is assumed to be linear.

Due to the limited response categories and also limited covariate subpopulations, assumptions (a) and (c) could not be checked visually using a scatter diagram. Also, one has to assume that the variability of response observations was constant, although due to the nature of the data, this was unlikely. This illustrates some of the problems encountered when one tries to apply linear regression method to ordered categorical data.

(b) Model fitting

The linear regression models which were fitted included: (i) model with 'smoke' on its own; (ii) model with 'smoke' adjusting for 'heart attack' and (iii) model with both covariates and a 1st order interaction term. The *t-test statistic* (H_0: parameter = 0) was used to assess the significance of each term in the model on 1 df. The parameter estimates and the fitted mean values for the single effects and the two-covariate case models were obtained for each sub-population.

5.2.2 Binary Logistic Regression Models

The purpose of fitting several binary logistic models to the ordinal quality of life scale was to examine the similarities/differences in these models. From this, one could determine whether it was necessary to have several binary logistic models to describe the relationship of the outcome with the covariates or whether one model, regardless of which one, was ample. As these models were treated independently, the need for multiple testing did not arise.

There were four levels for the response and therefore there were three cut-points that divided the health status categories into binary groupings: 1st cut-point -'*excellent*' v. ('*good*', '*fair*', '*poor*'), 2nd cut-point- ('*excellent*', '*good*') v. ('*fair*', '*poor*') and 3rd cut-point- ('*excellent*', '*good*', '*fair*') v. '*poor*'. The data for the response categories were amalgamated to form the two binary categories with the first grouping coded as '1' and the other coded as '0'. The odds were based on grouping '0' compared to grouping '1' in order to keep in consistent with the ordinal regression models.

Three separate binary analyses were carried out using the maximum likelihood method and this is described in Appendix II (section 2). Each model was fitted with the unadjusted 'smoke' and then 'heart attack' was added into the model to assess the adjusted 'smoke' effect. The 1st order interaction term for each model was also included and a test based on the change in the –2 log-likelihood of the main effects model and the one including the interaction term was carried out (deviance statistic). If the change in the –2 log-likelihood was significant, the interaction term was included. The maximum likelihood parameter estimates with their standard errors were given and from these the odds ratios and their 95% confidence intervals were derived.

5.2.3 Ordinal Regression Models

The starting point for fitting ordinal regression models is to compute the observed sample marginal probabilities (p_{ij}) and the observed sample response functions $F(p)$ together with their covariance-variance matrices (see Appendix II section 2 for detail). Once these have been obtained, the modelling stage can begin using either the method of maximum likelihood or weighted least squares.

(i) Maximum likelihood estimation

The parameter estimates (and their standard errors), the predicted response functions (and their variance covariance matrix) and the –2log-likelihood value of each of the following models were obtained using maximum likelihood estimation method.

- Polytomous model;
- Continuation ratio model (using separate binary logistic regressions);
- Stereotype model (with known constraints).

For these models the logits are of a generalised form and the log-likelihood functions are concave and the parameter estimates necessarily exist and are unique and finite if all the observed cell counts are positive. The mechanics of the maximum likelihood methods when used in the context of the above models are detailed in the Appendix II (section 2). In brief, the log-likelihood was computed and from this the first derivative with respect to the estimated parameters was obtained. This latter function comprised of the probabilities and the design matrix. The expected value of the second derivative was also computed and the Newton-Raphson method applied to obtain the maximum likelihood estimates of the parameters in the model.

(ii) Weighted least squares estimation

The parameter estimates (and their standard errors) and the predicted response functions (and their variance covariance matrix) of the following models were obtained using weighted least squares method. For these models where the ordinality of the categories was accounted for through the logits, weighted least squares estimation was more easily adaptable than maximum likelihood estimation. This was because the logits are a complex function of the cell counts and the likelihood function was not necessarily of a closed-form. This implies that the first and second order derivatives of the likelihood function are usually more difficult to obtain. Although maximum likelihood can be computed for the constant slope models (proportional odds, continuation ratio and the adjacent category), it was not decided to use the weighted least square method as the results could then be compared with other similar models. Hence, the weighted least squares estimation was used for the following:

- Proportional odds Model;
- Unconstrained Partial Proportional Odds Model;
- Constrained Partial Proportional Odds Model;
- Continuation Ratio (Unconstrained and fully constrained);
- Adjacent Category (constant and different slopes).

The weighted least squares estimate of the parameters is the vector that minimises the quadratic form (A12) as specified in Appendix II.

Once the maximum likelihood/weighted least square parameter estimates had been computed, questions about these parameters could be addressed using hypothesis testing (as in (A16) in Appendix II) and Wald test statistic ((A17) in Appendix II).

Below the observed sample marginal probabilities have been derived and these are common to all the ordinal regression models. Also, the observed sample response functions have been computed for the polytomous model and the cumulative logit model. The purpose of computing these by hand is to illustrate how the method using first principle works (using either maximum likelihood or weighted least squares). In practice one would not perform any calculations by hand, as the software package provides all the relevant information in the output (see Appendix II: section 3). The required design matrix has to be input by the user for modelling purposes and details of how the design matrix is determined are given below.

5.2.3.1 Polytomous Model

The observed sample response functions for the polytomous model are defined as:

$\ln\left(\frac{\Pr(Y_i = y_2)}{\Pr(Y_i = y_1)}\right)$ where this is the log odds based on categories '*good*' v. '*excellent*' and can be denoted by $f_{i1}p)$;

$\ln\left(\frac{\Pr(Y_i = y_3)}{\Pr(Y_i = y_1)}\right)$ is the log odds based on categories '*fair*' v. '*excellent*' and can be denoted by $f_{i2}(p)$;

$\ln\left(\frac{\Pr(Y_i = y_4)}{\Pr(Y_i = y_1)}\right)$ is the log odds based on categories '*poor*' v. '*excellent*' and can be denoted by $f_{i3}(p)$.

For the Health Status data, the polytomous model from (3.7) was of the form:

$$F(\pi) = \alpha_j + \sum_{k=1}^{2} x_{ik}\beta_{jk} \qquad j=1,..\,3. \qquad (5.1)$$

The '*excellent*' category was taken as referent.

The observed sample marginal probabilities and response functions for the polytomous model were computed as follows (see Appendix II (section 2) for details of the theory).

(a) Observed sample marginal probabilities

Given the Health Status data the observed sample marginal probabilities for each sub-population were:

p_{11}= 0.1111	p_{12}=0.3128	p_{13}=0.4156	p_{14}=0.1605
p_{21}=0.0797	P_{22}=0.3885	p_{23}=0.4230	p_{24}=0.1091
p_{31}=0.1897	p_{32}=0.4955	p_{33}=0.2463	p_{34}=0.0684
p_{41}=0.2146	p_{42}=0.4953	p_{43}=0.2457	p_{44}=0.0444.

Here

p_1'= (0.1111, 0.3128, 0.4156, 0.1605); p_2'= (0.0797, 0.3885, 0.4230, 0.1091)

p_3'= (0.1897, 0.4955, 0.2463, 0.0684); p_4'= (0.2146, 0.4953, 0.2457, 0.0444)

and p_1', p_2', p_3' and p_4' were each of (1x4) dimension and p'= (p_1', p_2', p_3' p_4') was of dimension (1x16).

The variance covariance matrix for p_1 was

$$V(p_1)=\begin{bmatrix} 1.8x10^{-5} & -1.2x10^{-5} & -0.6x10^{-5} & -0.1x10^{-5} \\ -1.2x10^{-5} & 2.7x10^{-5} & -1.3x10^{-5} & -0.2x10^{-5} \\ -0.6x10^{-5} & -1.3x10^{-5} & 2.0x10^{-5} & -0.1x10^{-5} \\ -0.1x10^{-5} & -0.2x10^{-5} & -0.1x10^{-5} & 4.24x10^{-2} \end{bmatrix}$$

and in a similar way $V(p_2)$, $V(p_3)$ and $V(p_4)$ were obtained and were each 4 x 4 in dimension.

Thus $V(p) = (V(p_1), V(p_2), V(p_3), V(p_4))$ and $V(p)$ had a dimension of 16 x 16.

SAS

The marginal and observed cell probabilities could be obtained in *SAS* as detailed in Appendix II (Section 3).

(b) Observed response functions

Then $f_{i1}(p) = ln\ (p_{i2}/p_{i1})$; $f_{i2}(p) = ln\ (p_{i3}/p_{i1})$; $f_{i3}(p) = ln\ (p_{i4}/p_{i1})$ and we can write $F(p_1) = \{f_{11}(p), f_{12}(p), f_{13}(p)\}$ and $F(p) = \{F(p_1), F(p_2), F(p_3), F(p_4)\}$.

We compute $F(p_1) = Kln(A(p_1))$, such that:

taking $\text{In}(Ap_1)=\text{In}\begin{bmatrix}1&0&0&0\\0&0&0&1\\0&1&0&0\\0&0&0&1\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}0.1111\\0.3128\\0.4156\\0.1605\end{bmatrix}$

(6 x 1) = (6 x 4) (4 x 1)

$$=\text{In}\begin{bmatrix}0.3128\\0.1111\\0.4156\\0.1111\\0.1605\\0.1111\end{bmatrix}=\begin{bmatrix}-1.1622\\-2.1973\\-0.8780\\-2.1973\\-1.8295\\-2.1973\end{bmatrix}$$

$$K = \begin{bmatrix}1&-1&&&&\\0&0&1&-1&&\\0&0&0&0&1&-1\end{bmatrix}$$

(3 x 6)

and

$$\text{F}(p_1) = K(\ln Ap_1) = \begin{bmatrix}1&-1&&&&\\0&0&1&-1&&\\0&0&0&0&1&-1\end{bmatrix}\begin{bmatrix}-1.1622\\-2.1973\\-0.8780\\-2.1973\\-1.8295\\-2.1973\end{bmatrix}=\begin{bmatrix}1.0351\\1.3193\\0.3679\end{bmatrix}.$$

(3x1)

Computing $F(p_2)$, $F(p_3)$, $F(p_4)$ in a similar way resulted in

$$F(p) = \{(1.0351, 1.3193, 0.3679); (1.5875, 1.6725, 0.3174); (0.9601, 0.2612, -1.0198); (0.8363, 0.1354, -1.5763)\}.$$

SAS

The calculation of the S matrix (see (A5) in Appendix II) was somewhat complex and laborious by hand, and therefore has not been detailed here. However, the observed response functions and the variance-covariance matrix S can be obtained in *SAS* as detailed in Appendix II (Section 3).

The model fitting could now be carried out.

(c) Model specification

To determine the parameter estimates, in addition to the probabilities, the design matrix X is required. For this, let us assume that an individual in group i takes on the values for 'smoke' g=1 (yes) or 2 (no) and 'heart attack' h=1 (yes) or 2 (no). Then for modelling purposes we write the response functions as:

$$f_{i1}(\hat{\pi}) = f_{gh1}(\hat{\pi}); \qquad f_{i2}(\hat{\pi}) = f_{gh2}(\hat{\pi}); \qquad f_{i3}(\hat{\pi}) = f_{gh3}(\hat{\pi}).$$

Instead of estimating one set of parameters for one logit function, as in the logistic regression model, one is estimating sets of parameters for multiple logit functions. This poses no particular problems, since there are multiple response functions being modelled per group, there are more degrees of freedom associated with each effect.

For the Health Status data, model (5.1) is expressed as:

$$\begin{bmatrix} f_{111}(\hat{\pi}) \\ f_{112}(\hat{\pi}) \\ f_{113}(\hat{\pi}) \\ f_{121}(\hat{\pi}) \\ f_{122}(\hat{\pi}) \\ f_{123}(\hat{\pi}) \\ f_{211}(\hat{\pi}) \\ f_{212}(\hat{\pi}) \\ f_{213}(\hat{\pi}) \\ f_{221}(\hat{\pi}) \\ f_{222}(\hat{\pi}) \\ f_{223}(\hat{\pi}) \end{bmatrix} = \begin{bmatrix} \alpha_1 + \beta_{11} + \beta_{12} \\ \alpha_2 + \beta_{21} + \beta_{22} \\ \alpha_3 + \beta_{31} + \beta_{32} \\ \alpha_1 + \beta_{11} - \beta_{12} \\ \alpha_2 + \beta_{21} - \beta_{22} \\ \alpha_3 + \beta_{31} - \beta_{32} \\ \alpha_1 - \beta_{11} + \beta_{12} \\ \alpha_2 - \beta_{21} + \beta_{22} \\ \alpha_3 - \beta_{31} + \beta_{32} \\ \alpha_1 - \beta_{11} - \beta_{12} \\ \alpha_2 - \beta_{21} - \beta_{22} \\ \alpha_3 - \beta_{31} - \beta_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \beta_{12} \\ \beta_{22} \\ \beta_{32} \end{bmatrix}. \quad (5.2)$$

The matrix formulation (5.2) contains the *design matrix* (also known as the *model matrix*) that consists of 0, -1 and 1 and the required parameters. Nine parameters were fitted for the main effects polytomous model using the Health Status data. Here the $\{\alpha_j\}$ were the average log odds (across the four sub-populations) of the health status effect and in terms of model (5.1) they were the intercept parameters for each cut-point specific logit function. These were nuisance parameters and therefore were of no interest when summarising the data. The $\{\beta_{j1}\}$ were the differential change in the log odds for smoking (yes/no) and $\{\beta_{j2}\}$ were the differential change in the log odds for having/not having suffered from a heart attack. For each cut-point $\{\beta_{j1}\}$ was the added amount for smokers and subtracted amount for non-smokers and $\{\beta_{j2}\}$ was the added amount for having had a heart attack and subtracted amount for not having had a heart attack. Table 5.1 displays the log odds predicted by this model.

Table 5.1: Table of model predicted log odds for the Polytomous Model using the Health Status data

Do you smoke? (g)	***Have you had a heart attack? (h)***	$f_{gh1}(\hat{\pi})$	$f_{gh2}(\hat{\pi})$	$f_{gh3}(\hat{\pi})$
Yes	***Yes***	$\alpha_1 + \beta_{11} + \beta_{12}$	$\alpha_2 + \beta_{21} + \beta_{22}$	$\alpha_3 + \beta_{31} + \beta_{32}$
Yes	***No***	$\alpha_1 + \beta_{11} - \beta_{12}$	$\alpha_2 + \beta_{21} - \beta_{22}$	$\alpha_3 + \beta_{31} - \beta_{32}$
No	***Yes***	$\alpha_1 - \beta_{11} + \beta_{12}$	$\alpha_2 - \beta_{21} + \beta_{22}$	$\alpha_3 - \beta_{31} + \beta_{32}$
No	***No***	$\alpha_1 - \beta_{11} - \beta_{12}$	$\alpha_2 - \beta_{21} - \beta_{22}$	$\alpha_3 - \beta_{31} - \beta_{32}$

(d) Estimation – parameter estimates and odds ratios

Maximum likelihood estimation was used to derive the parameter estimates and their standard errors by solving equation (A11) in Appendix II. This equation involved functions that comprised of the sample probabilities and the specified design matrix (in (5.2)).

Consequently, the predicted logits functions $F(\hat{\pi})$and their variance covariance matrix were computed. The log odds ratios and their 95% confidence intervals could be obtained in one of two ways:

(i) the parameter estimates could be substituted into the logit functions as given in Table 5.1 and as a by-product the log odds were provided. Alternatively, the logit functions as given in the *SAS* output could be used to compute the log odds. Subtracting an appropriate pair of logits, resulted in the log odds ratios and consequently the odds ratio ($e^{(\ln O.R)}$). For instance, to determine the adjusted log odds ratio of those with '*poor*' as opposed to '*excellent*' health status given smokers (as opposed to non-smokers) one would compute $(\alpha_1 + \beta_{11} + \beta_{12}) - (\alpha_1 - \beta_{11} + \beta_{12}) = 2\beta_{11}$. This method does not easily provide the standard error of the log odds ratios, and therefore the 95% confidence intervals are more difficult to obtain.

(ii) Another method for obtaining the log odds ratios involved the use of the contrast matrix that contained the linear combination of the parameter estimates. For instance, to obtain the log odds ratio and the standard error for the first cut-point for those who smoked as opposed to the non-smokers, the hypothesis H_0: $2\beta_{11}=0$ was set up and the contrast matrix C= [0 0 0 2 0 0 0 0 0] could be used to test this hypothesis - see Appendix II (A16) and (A17)) for details. The by-product of this was the log odds ratios and their standard errors. The odds ratios were then $e^{(\ln O.R)}$ and using the standard error of the log odds ratios, the 95% confidence interval for the odd ratios were computed as $e^{[\ln O.R \pm 1.96 (s.e. \ln O.R)]}$. Note that the standard error of In(O.R) is computed using the delta method.

Method (ii) was used to derive the odds ratios and their 95% confidence intervals for all the ordinal regression models fitted.

(e) Interaction term

The 1st order interaction term was constructed by taking the product of the main effects. This term was then incorporated into the main effects design matrix (in (5.2)) and the resulting design matrix, as in (5.3), was used to compute the parameter estimates and their variance covariance matrix (by the method of maximum likelihood) in very much the same way as above. The following matrix formulation provided the predicted logits $F(\hat{\pi})$:

$$\begin{bmatrix} f_{111}(\hat{\pi}) \\ f_{112}(\hat{\pi}) \\ f_{113}(\hat{\pi}) \\ f_{121}(\hat{\pi}) \\ f_{122}(\hat{\pi}) \\ f_{123}(\hat{\pi}) \\ f_{211}(\hat{\pi}) \\ f_{212}(\hat{\pi}) \\ f_{213}(\hat{\pi}) \\ f_{221}(\hat{\pi}) \\ f_{222}(\hat{\pi}) \\ f_{223}(\hat{\pi}) \end{bmatrix} = \begin{bmatrix} \alpha_1 + \beta_{11} + \beta_{12} + \beta_{11}\beta_{12} \\ \alpha_2 + \beta_{21} + \beta_{22} + \beta_{21}\beta_{22} \\ \alpha_3 + \beta_{31} + \beta_{32} + \beta_{31}\beta_{32} \\ \alpha_1 + \beta_{11} - \beta_{12} - \beta_{11}\beta_{12} \\ \alpha_2 + \beta_{21} - \beta_{22} - \beta_{21}\beta_{22} \\ \alpha_3 + \beta_{31} - \beta_{32} - \beta_{31}\beta_{32} \\ \alpha_1 - \beta_{11} + \beta_{12} - \beta_{11}\beta_{12} \\ \alpha_2 - \beta_{21} + \beta_{22} - \beta_{21}\beta_{22} \\ \alpha_3 - \beta_{31} + \beta_{32} - \beta_{31}\beta_{32} \\ \alpha_1 - \beta_{11} - \beta_{12} + \beta_{11}\beta_{12} \\ \alpha_2 - \beta_{21} - \beta_{22} + \beta_{21}\beta_{22} \\ \alpha_3 - \beta_{31} - \beta_{32} + \beta_{31}\beta_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ \beta_{11}\beta_{12} \\ \beta_{21}\beta_{22} \\ \beta_{31}\beta_{32} \end{bmatrix}. \tag{5.3}$$

Here the interaction effects were $\beta_{11}\beta_{12}$, $\beta_{21}\beta_{22}$ and $\beta_{31}\beta_{32}$. In this saturated model 12 parameters were fitted equalling the total of the number of logits in the contingency table. The significance of the interaction term was based on assessing the change in the –2log-likelihoods of the two models (with and without the interaction term). This deviance statistic was distributed using 3-df (degrees of freedom), and this was the difference in the degrees of freedom based on the two models.

5.2.3.2 Cumulative Logit Models

The Health Status response scale does not have an underlying continuum. However, according to McCullagh (1980) the cumulative logit models can still be fitted; the only problem encountered is that the interpretation is not as easy as when presented with a *y*-response that has an underlying continuum.

For the proportional odds and the partial proportional odds models, weighted least squares method was used to obtain the parameter estimates and provide the statistical inference. The general principles of this method are detailed in Appendix II – Sections 2 and 3.

(a) Observed sample marginal probabilities

These were as specified for the polytomous model (see section 5.2.3.1 (a)).

(b) Observed sample response functions

The cumulative logits are expressed as in (3.8). In terms of the health status response categories the sample logits can be written as follows:

$\ln\left(\frac{\Pr(Y_i > y_1)}{\Pr(Y_i \le y_1)}\right)$ is the log odds for ('*good*', '*fair*', '*poor*') v. '*excellent*' and can be denoted by $f_{i1}(p)$;

$\ln\left(\frac{\Pr(Y_i > y_2)}{\Pr(Y_i \le y_2)}\right)$ is the log odds for ('*fair*', '*poor*') v. ('*excellent*', '*good*') and can be denoted by $f_{i2}(p)$;

$\ln\left(\frac{\Pr(Y_i > y_3)}{\Pr(Y_i \le y_3)}\right)$ is the log odds for '*poor*' v. ('*excellent*', '*good*', '*fair*') and can be denoted by $f_{i3}(p)$.

The observed sample response functions together with their standard errors are obtained in a similar way to those of the polytomous model. We usually define the cumulative probabilities for a cumulative logits model as:

$$\phi_{i1} = \pi_{i1};\ \ \phi_{i2} = \pi_{i1}+\pi_{i2};\ \ \phi_{i3} = \pi_{i1}+\pi_{i2}+\pi_{i3};\ \ \ldots\ldots\ \ \phi_{ic-1} = \pi_{i1}+\ldots\ldots+\pi_{ic-1};\ \ \phi_{ic} = 1.$$

However, to keep the parameters consistent with those of *SAS*, the cumulative probabilities were defined as:

$$\phi_{i1} = 1;\ \ \phi_{i2} = \pi_{i2}+\ldots\ldots+\pi_{ic};\ \ \phi_{i3} = \pi_{i3}+..+\pi_{ic};\ \ \ldots\ldots\ \ \phi_{ic} = \pi_{i1}+\ldots\ldots+\pi_{ic};\ \ \phi_{ic} = \pi_{ic}.$$

Using these latter probabilities, the sample cumulative logit response functions were:-

$f_{i1}(p) = ln(\phi_{i2}/(1-\phi_{i2}))$;

.

.

$f_{i3}(p) = ln\ (\phi_{i4}/(1-\phi_{i4}))$.

As before, we write $F(p_1) = \{f_{11}(p), f_{12}(p), f_{13}(p)\}$ and

$F(p) = \{F(p_1), F(p_2), F(p_3), F(p_4)\}$. Then we compute $F(p_1) = K\ ln\ (A(p_1))$, such that:

$$ln\ (A(p_1)) = \ln \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.1111 \\ 0.3128 \\ 0.4156 \\ 0.1605 \end{bmatrix}$$

and

$$F(p_1) = K(\ln(Ap_1)) = \begin{bmatrix} 1 & -1 & & & & \\ 0 & 0 & 1 & -1 & & \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} -0.1178 \\ -2.1972 \\ -0.5514 \\ -0.8583 \\ -1.8295 \\ -0.1749 \end{bmatrix} = \begin{bmatrix} 2.0794 \\ 0.3069 \\ -1.6546 \end{bmatrix}.$$

Computing $F(p_2)$, $F(p_3)$, $F(p_4)$ in a similar way results in $F(p)$. The variance of $F(p)$ was derived using the same method as specified for the polytomous model.

The design matrix, X, was required in addition to the observed sample marginal probabilities and logits, so as that modelling could be carried out. The design matrix varies depending on the type of cumulative model fitted. There are three types of cumulative models considered here: (i) different slopes cumulative logit model, (ii) proportional odds and (iii) partial proportional odds models.

(i) Different slopes Cumulative Logit Model

Prior to fitting the proportional odds model and the partial proportional odds models, we consider the cumulative logit model with cut-point specific parameters. The purpose of this model was to give an idea of how the odds ratios over the cut-points were behaving and which model, whether it be the proportional odds or the partial proportional odds, was appropriate for the data. Assessing the standard errors of the adjusted and unadjusted 'smoke' covariate did not arise, as the purpose of this different slopes cumulative model was purely to observe visually the individual odds ratios over the cut-points (no testing of the homogeneity of the cut-point specific parameters was carried out).

(a) Model specification

The general form of the different slopes cumulative model is given in (3.9). For the Health Status data this model for the response functions was specified as:

$$F(\pi) = \alpha_j + \sum_{k=1}^{2} x_{ik}\beta_{jk} \qquad j=1, 2, 3 \qquad (5.4)$$

The main effects different slopes cumulative logit model fitted separate slope parameters for each covariate and therefore the number of parameters was the same as for the main effects polytomous model. Thus, the design matrix used for the polytomous model (in (5.2)) was identical to the design matrix used for the different slopes cumulative model.

(b) Estimation – parameter estimates and odds ratios

The observe sample marginal probabilities and logits (as derived above), together with the design matrix (in (5.2)) were used to compute the weighted least square equation (A13) given in Appendix II. The by-product of this was the parameter estimates. The standard errors of these estimates were obtained using (A14) in Appendix II.

The predicted response functions, $F(\hat{\pi})$, were derived by solving equation (5.2) with the cumulative logit specification. Consequently, the cut-point specific log odds ratios were derived using method 5.2.3.1 (d).

(c) Interaction term

The interaction term for this model was not considered, as attention was primarily centred round the main effects and the behaviour of the log odds ratios.

(ii) Proportional Odds Model

The general form of the proportional odds model is given in (3.10). For the Health Status data, the model for the response functions was expressed as:

$$F(\pi) = \alpha_j + \sum_{k=1}^{2} x_{ik}\beta_k \qquad j=1, 2, 3 \qquad (5.5)$$

(a) Model specification

One effectively constrains all the cut-point specific regression coefficients of a covariate in the different slopes cumulative logit model to be the same when fitting the proportional odds model. As before if we assume that an individual in group i takes on the values for 'smoke' g=1 (yes) or 2 (no) and 'heart attack' h=1 (yes) or 2 (no), then for modelling purposes (5.5) can be specified using the following matrix formulation.

In this model five parameters were fitted. The $\{\alpha_j\}$ parameters were as defined for the different slopes cumulative model and were the nuisance parameters. The β_1 parameter was the differential change in the log odds for smoking (yes/no) and β_2 as the differential change in the log odds for suffering/not suffering from a heart attack. β_1 was the added amount for smoking and subtracted amount for non-smoking and β_2 was the added amount for having had a heart attack and subtracted amount for not having had a heart attack over all the cut-points. The predicted log odds for this model were as given in Table 5.2.

$$
\begin{bmatrix} f_{111}(\hat{\pi}) \\ f_{112}(\hat{\pi}) \\ f_{113}(\hat{\pi}) \\ f_{121}(\hat{\pi}) \\ f_{122}(\hat{\pi}) \\ f_{123}(\hat{\pi}) \\ f_{211}(\hat{\pi}) \\ f_{212}(\hat{\pi}) \\ f_{213}(\hat{\pi}) \\ f_{221}(\hat{\pi}) \\ f_{222}(\hat{\pi}) \\ f_{223}(\hat{\pi}) \end{bmatrix} = \begin{bmatrix} \alpha_1 + \beta_1 + \beta_2 \\ \alpha_2 + \beta_1 + \beta_2 \\ \alpha_3 + \beta_1 + \beta_2 \\ \alpha_1 + \beta_1 - \beta_2 \\ \alpha_2 + \beta_1 - \beta_2 \\ \alpha_3 + \beta_1 - \beta_2 \\ \alpha_1 - \beta_1 + \beta_2 \\ \alpha_2 - \beta_1 + \beta_2 \\ \alpha_3 - \beta_1 + \beta_2 \\ \alpha_1 - \beta_1 - \beta_2 \\ \alpha_2 - \beta_1 - \beta_2 \\ \alpha_3 - \beta_1 - \beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 & -1 \\ 1 & 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & 1 \\ 1 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{bmatrix} \qquad (5.6)
$$

Table 5.2: Table of model predicted log odds for the Proportional Odds Model using the Health Status data

Do you smoke? (g)	**Have you had a heart attack? (h)**	$f_{gh1}(\hat{\pi})$	$f_{gh2}(\hat{\pi})$	$f_{gh3}(\hat{\pi})$
Yes	**Yes**	$\alpha_1 + \beta_1 + \beta_2$	$\alpha_2 + \beta_1 + \beta_2$	$\alpha_3 + \beta_1 + \beta_2$
Yes	**No**	$\alpha_1 + \beta_1 - \beta_2$	$\alpha_2 + \beta_1 - \beta_2$	$\alpha_3 + \beta_1 - \beta_2$
No	**Yes**	$\alpha_1 - \beta_1 + \beta_2$	$\alpha_2 - \beta_1 + \beta_2$	$\alpha_3 - \beta_1 + \beta_2$
No	**No**	$\alpha_1 - \beta_1 - \beta_2$	$\alpha_2 - \beta_1 - \beta_2$	$\alpha_3 - \beta_1 - \beta_2$

(b) Estimation - parameter estimates and odds ratio

The design matrix in equation (5.6) together with the observed sample marginal probabilities and response functions (derived above) were used to solve equations (A13) and (A14) in Appendix II. The by-product of this was the parameter estimates and their standard errors.

The predicted logits, $F(\hat{\pi})$, were obtained by solving (5.6) and consequently the resultant was the odds ratios together with their 95% confidence intervals (as in 5.2.3.1(d) (ii)).

(c) Interaction term

The 1st order interaction term for the proportional odds model was obtained by the cross-product of the main effects. This term was included in the model and the following resulted:

$$\begin{bmatrix} f_{111}(\hat{\pi}) \\ f_{112}(\hat{\pi}) \\ f_{113}(\hat{\pi}) \\ f_{121}(\hat{\pi}) \\ f_{122}(\hat{\pi}) \\ f_{123}(\hat{\pi}) \\ f_{211}(\hat{\pi}) \\ f_{212}(\hat{\pi}) \\ f_{213}(\hat{\pi}) \\ f_{221}(\hat{\pi}) \\ f_{222}(\hat{\pi}) \\ f_{223}(\hat{\pi}) \end{bmatrix} = \begin{bmatrix} \alpha_1 + \beta_1 + \beta_2 + \beta_1\beta_2 \\ \alpha_2 + \beta_1 + \beta_2 + \beta_1\beta_2 \\ \alpha_3 + \beta_1 + \beta_2 + \beta_1\beta_2 \\ \alpha_1 + \beta_1 - \beta_2 - \beta_1\beta_2 \\ \alpha_2 + \beta_1 - \beta_2 - \beta_1\beta_2 \\ \alpha_3 + \beta_1 - \beta_2 - \beta_1\beta_2 \\ \alpha_1 - \beta_1 + \beta_2 - \beta_1\beta_2 \\ \alpha_2 - \beta_1 + \beta_2 - \beta_1\beta_2 \\ \alpha_3 - \beta_1 + \beta_2 - \beta_1\beta_2 \\ \alpha_1 - \beta_1 - \beta_2 + \beta_1\beta_2 \\ \alpha_2 - \beta_1 - \beta_2 + \beta_1\beta_2 \\ \alpha_3 - \beta_1 - \beta_2 + \beta_1\beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 \\ 0 & 1 & 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \\ 1 & 0 & 0 & -1 & 1 & -1 \\ 0 & 1 & 0 & -1 & 1 & -1 \\ 0 & 0 & 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & -1 & -1 & 1 \\ 0 & 1 & 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \beta_1\beta_2 \end{bmatrix} . \qquad (5.7)$$

The interaction term corresponded to the regression parameter $\beta_1\beta_2$ in (5.7). The parameter estimates and their standard errors were obtained as above using the weighted least square equations. The Wald test statistic was computed using the parameter estimate and the standard error of the interaction term and this test determined whether this term was significant or not (as in Appendix II (A17)). This was set up in a contrast statement and the test was based on 1-df (since there was one linearly independent row in the contrast matrix).

(d) Proportional Odds assumption

Various methods have been cited in section 3.5.2 (b) that permit the assumption of proportional odds to be tested. The likelihood ratio test, score test and the Wald test would be the most likely ones to be used here, as the other methods are not available in *SAS*. However, as the analysis was based on weighted least squares, the score test and Wald test could only be used. The need to use graphical methods did not arise, as difference in the cut-point specific regression parameters given a covariate, was on inspection, easily visible.

The chi-squared score test statistic produced in *PROC LOGISTIC* in *SAS* provided an overall assessment of the proportional odds for all the covariates. Where the assumption was violated, the non-proportionality was examined further for each individual covariate by using the Wald test statistic (as in Appendix II (A16) and (A17)).

For instance, in the different slopes cumulative model (5.4), the β vector consisted of

$$(\alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \beta_{11} \quad \beta_{21} \quad \beta_{31} \quad \beta_{12} \quad \beta_{22} \quad \beta_{32}). \quad (5.8)$$

The hypothesis set up to assess whether the differential effect of 'smoke' was similar over the three cut-point categories was H_0: $\beta_{11}=\beta_{21}=\beta_{31}$. This hypothesis was set up as H_0: $\beta_{11}-\beta_{21}=\beta_{21}-\beta_{31}=\beta_{11}-\beta_{31}=0$. It was tested using a contrast matrix, C in conjunction with Wald test statistic. The C matrix in the different slopes cumulative model was specified as:

$$C_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}. \quad (5.9)$$

A similar contrast statement was repeated to test whether differential effect of 'heart attack' was similar over the three cut-point categories. The C matrix was specified as:

$$C_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}. \quad (5.10)$$

The χ^2 degrees of freedom were the number of linearly independent rows in the contrast matrix: thus both tests were based on 2-df.

(iii) Unconstrained Partial Proportional Odds Model

There was evidence that the proportional odds assumption was violated and consequently alternative models such as the partial proportional odds models had to be considered.

(a) Model specification

For the partial proportional odds models (unconstrained and constrained) the systematic component of equation (3.14) consisted of some variables where constant odds could be assumed and some, where the regression parameters were allowed to vary over the cut-points. For the Health Status data, model (3.14) for the response functions took on the form

$$F(\pi) = \alpha_j + x_{i1}\beta_1 + x_{i2}\beta_2 + T_{i1}\gamma_{j1} \quad \text{where } j=1,2,3. \qquad (5.11)$$

This model could be specified using the following matrix formulation (where the *i* group has *gh* sub-population):

$$\begin{bmatrix} f_{111}(\hat{\pi}) \\ f_{112}(\hat{\pi}) \\ f_{113}(\hat{\pi}) \\ f_{121}(\hat{\pi}) \\ f_{122}(\hat{\pi}) \\ f_{123}(\hat{\pi}) \\ f_{211}(\hat{\pi}) \\ f_{212}(\hat{\pi}) \\ f_{213}(\hat{\pi}) \\ f_{221}(\hat{\pi}) \\ f_{222}(\hat{\pi}) \\ f_{223}(\hat{\pi}) \end{bmatrix} = \begin{bmatrix} \alpha_1 + (\beta_1 + \gamma_{11}) + \beta_2 \\ \alpha_2 + (\beta_1 + \gamma_{21}) + \beta_2 \\ \alpha_3 + (\beta_1 + \gamma_{31}) + \beta_2 \\ \alpha_1 + (\beta_1 + \gamma_{11}) - \beta_2 \\ \alpha_2 + (\beta_1 + \gamma_{21}) - \beta_2 \\ \alpha_3 + (\beta_1 + \gamma_{31}) - \beta_2 \\ \alpha_1 - (\beta_1 + \gamma_{11}) + \beta_2 \\ \alpha_2 - (\beta_1 + \gamma_{21}) + \beta_2 \\ \alpha_3 - (\beta_1 + \gamma_{31}) + \beta_2 \\ \alpha_1 - (\beta_1 + \gamma_{11}) - \beta_2 \\ \alpha_2 - (\beta_1 + \gamma_{21}) - \beta_2 \\ \alpha_3 - (\beta_1 + \gamma_{31}) - \beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \\ \beta_2 \end{bmatrix} . \quad 5.12$$

In this model 7 parameters were fitted (since γ_{11}=0). The interpretation of $\{\alpha_j\}$ and $\{\beta_k\}$ parameters was similar to that in the proportional odds model and the $\{\gamma_j\}$s were the differential change in the log odds of smoking (yes/no) associated with each jth cumulative logit. The γ_{11}=0 as the first logit was based on β_1 and this was incorporated into the design matrix.

The predicted logits were of the form given in Table 5.3.

Table 5.3: Table of model predicted log odds for the Unconstrained Partial Proportional Odds Model using the Health Status data

Do you smoke ? (g)	***Have you had a heart attack? (h)***	$f_{gh1}(\hat{\pi})$	$f_{gh2}(\hat{\pi})$	$f_{gh3}(\hat{\pi})$
Yes	***Yes***	$\alpha_1 + (\beta_1 + \gamma_{11}) + \beta_2$	$\alpha_2 + (\beta_1 + \gamma_{21}) + \beta_2$	$\alpha_3 + (\beta_1 + \gamma_{31}) + \beta_2$
Yes	***No***	$\alpha_1 + (\beta_1 + \gamma_{11}) - \beta_2$	$\alpha_2 + (\beta_1 + \gamma_{21}) - \beta_2$	$\alpha_3 + (\beta_1 + \gamma_{31}) - \beta_2$
No	***Yes***	$\alpha_1 - (\beta_1 + \gamma_{11}) + \beta_2$	$\alpha_2 - (\beta_1 + \gamma_{21}) + \beta_2$	$\alpha_3 - (\beta_1 + \gamma_{31}) + \beta_2$
No	***No***	$\alpha_1 - (\beta_1 + \gamma_{11}) - \beta_2$	$\alpha_2 - (\beta_1 + \gamma_{21}) - \beta_2$	$\alpha_3 - (\beta_1 + \gamma_{31}) - \beta_2$

(b) Estimation - parameter estimates and odds ratio

The parameter estimates were obtained using weighted least squares equations (A13) and (A14) in Appendix II. To solve these equations, the design matrix, as given in (5.12), the observed sample marginal probabilities and the sample response functions were used (as obtained above).

The predicted logits $F(\hat{\pi})$ were obtained by solving equation (5.12) and consequently the log odds ratios resulted. The constant log odds ratio for those who had suffered/not suffered a heart attack were obtained together with the cut-point specific log odds ratios for those who smoked/did not smoke by the linear combination of the different parameters. For instance, for the different cut-points, the log odds ratio of obtaining ('*good*', '*fair*', '*poor*') as opposed to '*excellent*' health status for the smokers (against the non-smokers) was $[(\alpha_1 + (\beta_1 + \gamma_{11}) + \beta_2) - (\alpha_1 - (\beta_1 + \gamma_{11}) + \beta_2)] = 2(\beta_1 + \gamma_{11}) = 2\beta_1$ (since $\gamma_{11} = 0$). Likewise the log odds ratio of obtaining ('*fair*', '*poor*') as opposed to ('*excellent*', '*good*') health status for the smokers versus non-smokers was $2(\beta_1 + \gamma_{21})$ and '*poor*' as opposed to ('*excellent*', '*good*', '*fair*') health status for the smokers versus non-smokers was $2(\beta_1 + \gamma_{31})$. The adjusted log odds ratio for those who suffered from a heart attack, as opposed to not having suffered from a heart attack was the same as that obtained for the proportional odds model. These log odds ratios were obtained using the contrast statements and the resultant was the odds ratios and their 95% confidence intervals.

<u>(c) Interaction term</u>

The cross products of the main effects were constructed resulting in the effects for the interaction term. The interaction term was made up of two components: $\beta_1\beta_2 + \beta_2\gamma_{j1}$. These effects were incorporated into the design matrix in equation (5.12) and the resulted in:

$$
\begin{bmatrix} f_{111}(\hat{\pi}) \\ f_{112}(\hat{\pi}) \\ f_{113}(\hat{\pi}) \\ f_{121}(\hat{\pi}) \\ f_{122}(\hat{\pi}) \\ f_{123}(\hat{\pi}) \\ f_{211}(\hat{\pi}) \\ f_{212}(\hat{\pi}) \\ f_{213}(\hat{\pi}) \\ f_{221}(\hat{\pi}) \\ f_{222}(\hat{\pi}) \\ f_{223}(\hat{\pi}) \end{bmatrix} = \begin{bmatrix} \alpha_1 + (\beta_1 + \gamma_{11}) + \beta_2 + \beta_1\beta_2 + \beta_2\gamma_{11} \\ \alpha_2 + (\beta_1 + \gamma_{21}) + \beta_2 + \beta_1\beta_2 + \beta_2\gamma_{21} \\ \alpha_3 + (\beta_1 + \gamma_{31}) + \beta_2 + \beta_1\beta_2 + \beta_2\gamma_{31} \\ \alpha_1 + (\beta_1 + \gamma_{11}) - \beta_2 - \beta_1\beta_2 - \beta_2\gamma_{11} \\ \alpha_2 + (\beta_1 + \gamma_{21}) - \beta_2 - \beta_1\beta_2 - \beta_2\gamma_{21} \\ \alpha_3 + (\beta_1 + \gamma_{31}) - \beta_2 - \beta_1\beta_2 - \beta_2\gamma_{31} \\ \alpha_1 - (\beta_1 + \gamma_{11}) + \beta_2 - \beta_1\beta_2 - \beta_2\gamma_{11} \\ \alpha_2 - (\beta_1 + \gamma_{21}) + \beta_2 - \beta_1\beta_2 - \beta_2\gamma_{21} \\ \alpha_3 - (\beta_1 + \gamma_{31}) + \beta_2 - \beta_1\beta_2 - \beta_2\gamma_{31} \\ \alpha_1 - (\beta_1 + \gamma_{11}) - \beta_2 + \beta_1\beta_2 + \beta_2\gamma_{11} \\ \alpha_2 - (\beta_1 + \gamma_{21}) - \beta_2 + \beta_1\beta_2 + \beta_2\gamma_{21} \\ \alpha_3 - (\beta_1 + \gamma_{31}) - \beta_2 + \beta_1\beta_2 + \beta_2\gamma_{31} \end{bmatrix} =
$$

$$
\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & -1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \\ \beta_2 \\ \beta_1\beta_2 \\ \beta_2\gamma_{11} \\ \beta_2\gamma_{21} \\ \beta_2\gamma_{31} \end{bmatrix} . \qquad (5.13)
$$

The design matrix in (5.13) together with the observed sample marginal probabilities and logits were used to solve the weighted least square equations as detailed in Appendix II and the end-product was the parameter estimates.

The hypotheses H_{01}: $\beta_1\beta_2=0$ and H_{02}: $\beta_2\gamma_{21}=\beta_2\gamma_{31}=0$ (since $\gamma_{11}=0$) were constructed to test the significance of the interaction term. The following contrast matrices were used:

$$C_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \qquad (5.14)$$

and

$$C_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (5.15)$$

Note, that the 1st row of matrix C_2 consisted of all zeros due to $\gamma_{11}=0$. The Wald test statistic was used to assess the significance of the latter effects in the model (see Appendix II (A16) and (A17)). The H_{01} test was based on 1-df and H_{02} was based on 2-df. (i.e. the number of linearly independent rows in the contrast matrix).

(d) Fitting the unadjusted 'smoke' covariate

The unadjusted 'smoke' covariate could not be fitted, as this was made up of two regression components, β and γ. Also, the number of parameters required out numbered the number of logits (i.e. there was over-parameterisation) in the design matrix. Thus the Asymptotic Relative Precision (ARP) could not be computed for this model.

(iv) Constrained Partial Proportional Odds Model

(a) Model specification

The cut-point specific log odds ratios were observed in the unconstrained partial proportional odds model. A monotonic trend was apparent in the log odds ratios across the health status categories in relation to the smokers/non-smokers. To simplify the interpretation, a constraint was placed on the regression parameters associated with the 'smoke' covariate (leading to the formation of the constrained partial proportional odds model). The 'heart attack' covariate was assumed to have constant log odds over all the cut-points. The constrained partial proportional odds model of the form (3.16) was fitted. For the Health Status data, this modelled the predicted response functions using:

$$F(\pi) = \alpha_j + x_{i1}\beta_1 + x_{i2}\beta_2 + T_{i1}\Gamma_j\gamma_1 \qquad \text{where } j=1, 2, 3. \qquad (5.16)$$

Since $\gamma_{11}=0$ in the unconstrained model, then $\Gamma_1=0$ in model (5.16). Thus the first logit was based on the β_1. The logits based on the cut-points 2 and 3 were defined as $\beta_1+\Gamma_2\gamma_1$ and $\beta_1+\Gamma_3\gamma_1$ respectively. In the unconstrained partial proportional odds model γ_{21} and γ_{31} have the values 0.00411 and 0.1691 respectively, and from these one could derive that $\gamma_{31}=40\gamma_{21}$ approximately. The following constraints were chosen: $\Gamma_1=0$; $\Gamma_2=1$; $\Gamma_3=40$ and model (5.16) was fitted using the matrix formulation (where the i group has gh sub-population) given below.

$$\begin{bmatrix} f_{111}(\hat{\pi}) \\ f_{112}(\hat{\pi}) \\ f_{113}(\hat{\pi}) \\ f_{121}(\hat{\pi}) \\ f_{122}(\hat{\pi}) \\ f_{123}(\hat{\pi}) \\ f_{211}(\hat{\pi}) \\ f_{212}(\hat{\pi}) \\ f_{213}(\hat{\pi}) \\ f_{221}(\hat{\pi}) \\ f_{222}(\hat{\pi}) \\ f_{223}(\hat{\pi}) \end{bmatrix} = \begin{bmatrix} \alpha_1 + (\beta_1 + 0\gamma_1) + \beta_2 \\ \alpha_2 + (\beta_1 + 1\gamma_1) + \beta_2 \\ \alpha_3 + (\beta_1 + 40\gamma_1) + \beta_2 \\ \alpha_1 + (\beta_1 + 0\gamma_1) - \beta_2 \\ \alpha_2 + (\beta_1 + 1\gamma_1) - \beta_2 \\ \alpha_3 + (\beta_1 + 40\gamma_1) - \beta_2 \\ \alpha_1 - (\beta_1 + 0\gamma_1) + \beta_2 \\ \alpha_2 - (\beta_1 + 1\gamma_1) + \beta_2 \\ \alpha_3 - (\beta_1 + 40\gamma_1) + \beta_2 \\ \alpha_1 - (\beta_1 + 0\gamma_1) - \beta_2 \\ \alpha_2 - (\beta_1 + 1\gamma_1) - \beta_2 \\ \alpha_3 - (\beta_1 + 40\gamma_1) - \beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 40 & 1 \\ 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 1 & 40 & -1 \\ 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & -1 & -40 & 1 \\ 1 & 0 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & -1 & -40 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \gamma_1 \\ \beta_2 \end{bmatrix} \qquad (5.17)$$

The predicted log odds using this specification are as detailed in Table 5.4.

Table 5.4: Table of model predicted log odds for the Constrained Partial Proportional odds Model using the Health Status data

Do you smoke? ***(g)***	***Have you had a heart attack?*** ***(h)***	$f_{gh1}(\hat{\pi})$	$f_{gh2}(\hat{\pi})$	$f_{gh3}(\hat{\pi})$
Yes	***Yes***	$\alpha_1 + (\beta_1 + 0\gamma_1) + \beta_2$	$\alpha_2 + (\beta_1 + 1\gamma_1) + \beta_2$	$\alpha_3 + (\beta_1 + 40\gamma_1) + \beta_2$
Yes	***No***	$\alpha_1 + (\beta_1 + 0\gamma_1) - \beta_2$	$\alpha_2 + (\beta_1 + 1\gamma_1) - \beta_2$	$\alpha_3 + (\beta_1 + 40\gamma_1) - \beta_2$
No	***Yes***	$\alpha_1 - (\beta_1 + 0\gamma_1) + \beta_2$	$\alpha_2 - (\beta_1 + 1\gamma_1) + \beta_2$	$\alpha_3 - (\beta_1 + 40\gamma_1) + \beta_2$
No	***No***	$\alpha_1 - (\beta_1 + 0\gamma_1) - \beta_2$	$\alpha_2 - (\beta_1 + 1\gamma_1) - \beta_2$	$\alpha_3 - (\beta_1 + 40\gamma_1) - \beta_2$

Six parameters were fitted for the constrained partial proportional odds model and β_1 and β_2 were the differential change in the log odds over all the cut-points for each covariate respectively. The γ_1 was the added differential change in the log odds for the 'smoke' covariate, with the Γ_j scalars (Γ_1=0; Γ_2=1; Γ_3=40) associated with each of the j^{th} logits.

(b) Estimation - parameter estimates and odds ratio

The design matrix in equation (5.17) together with the observed sample marginal probabilities and response functions were used to solve the weighted least square equations (as detailed in Appendix II). This resulted in the parameter estimates.

The predicted response functions $F(\hat{\pi})$ were obtained from (5.17) and consequently resulted in the log odds. The constant log odds ratio of some form of 'worse' health status compared to 'better' health for those who had suffered from a heart attack against those who have not had a heart attack was $2\beta_2$. The log odds ratio (*'good', 'fair', 'poor'*) against *'excellent'* health status for those who smoked (against those who did not smoke) was $[\alpha_1 + (\beta_1 + 0\gamma_1) + \beta_2]$-$[\alpha_1 - (\beta_1 + 0\gamma_1) + \beta_2] = 2\beta_1$. The log odds ratio (*'poor', 'fair'*) against (*'excellent', 'good'*) health status for those who smoked (against those who did not smoke) was $[\alpha_2 + (\beta_1 + 1\gamma_1) + \beta_2]$-$[\alpha_2 - (\beta_1 + 1\gamma_1) + \beta_2] = 2(\beta_1 + \gamma_1)$ and the log odds ratio for *'poor'* versus (*'excellent', 'good', 'fair'*) health was $[\alpha_2 + (\beta_1 + 40\gamma_1) + \beta_2]$-$[\alpha_2 - (\beta_1 + 40\gamma_1) + \beta_2] = 2(\beta_1 + 40\gamma_1)$. These log odds ratios and the odds ratios with their standard errors were direct product of specifying and solving the hypotheses: H_{01}: $2\beta_1=0$; H_{02}: $2(\beta_1+\gamma_1)=0$; H_{03}: $2(\beta_1+40\gamma_1)=0$ and H_{04}: $2\beta_2=0$. The contrast matrix for each hypothesis respectively was:

$$C_1=[0 \quad 0 \quad 0 \quad 2 \quad 0 \quad 0]; \tag{5.18}$$

$$C_2=[0 \quad 0 \quad 0 \quad 2 \quad 2 \quad 0]; \tag{5.19}$$

$$C_3=[0 \quad 0 \quad 0 \quad 2 \quad 80 \quad 0]; \tag{5.20}$$

$$C_4=[0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 2]. \tag{5.21}$$

Note, that the purpose of these contrasts was purely to obtain the estimates of the odds ratios and their confidence intervals as in 5.2.3.1 (d) (ii) (no testing was carried out).

(c) Interaction term

The interaction term for this model was constructed in a similar way to that of the unconstrained partial proportional odds model. The constrained partial proportional odds model with an interaction term took on the form:

$$\begin{bmatrix} f_{111}(\hat{\pi}) \\ f_{112}(\hat{\pi}) \\ f_{113}(\hat{\pi}) \\ f_{121}(\hat{\pi}) \\ f_{122}(\hat{\pi}) \\ f_{123}(\hat{\pi}) \\ f_{211}(\hat{\pi}) \\ f_{212}(\hat{\pi}) \\ f_{213}(\hat{\pi}) \\ f_{221}(\hat{\pi}) \\ f_{222}(\hat{\pi}) \\ f_{223}(\hat{\pi}) \end{bmatrix} = \begin{bmatrix} \alpha_1+(\beta_1+0\gamma_1)+\beta_2+(\beta_1\beta_2+0\gamma_1\beta_2) \\ \alpha_2+(\beta_1+1\gamma_1)+\beta_2+(\beta_1\beta_2+1\gamma_1\beta_2) \\ \alpha_3+(\beta_1+40\gamma_1)+\beta_2+(\beta_1\beta_2+40\gamma_1\beta_2) \\ \alpha_1+(\beta_1+0\gamma_1)-\beta_2-(\beta_1\beta_2+0\gamma_1\beta_2) \\ \alpha_2+(\beta_1+1\gamma_1)-\beta_2-(\beta_1\beta_2+1\gamma_1\beta_2) \\ \alpha_3+(\beta_1+40\gamma_1)-\beta_2-(\beta_1\beta_2+40\gamma_1\beta_2) \\ \alpha_1-(\beta_1+0\gamma_1)+\beta_2-(\beta_1\beta_2+0\gamma_1\beta_2) \\ \alpha_2-(\beta_1+1\gamma_1)+\beta_2-(\beta_1\beta_2+1\gamma_1\beta_2) \\ \alpha_3-(\beta_1+40\gamma_1)+\beta_2-(\beta_1\beta_2+40\gamma_1\beta_2) \\ \alpha_1-(\beta_1+0\gamma_1)-\beta_2+(\beta_1\beta_2+0\gamma_1\beta_2) \\ \alpha_2-(\beta_1+1\gamma_1)-\beta_2+(\beta_1\beta_2+1\gamma_1\beta_2) \\ \alpha_3-(\beta_1+40\gamma_1)-\beta_2+(\beta_1\beta_2+40\gamma_1\beta_2) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 40 & 1 & 1 & 40 \\ 1 & 0 & 0 & 1 & 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 1 & 40 & -1 & -1 & -40 \\ 1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 & -1 & 1 & -1 & -1 \\ 0 & 0 & 1 & -1 & -40 & 1 & -1 & -40 \\ 1 & 0 & 0 & -1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & -1 & -1 & 1 & 1 \\ 0 & 0 & 1 & -1 & -40 & -1 & 1 & 40 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \gamma_1 \\ \beta_2 \\ \beta_1\beta_2 \\ \beta_2\gamma_1 \end{bmatrix}$$

(5.22).

The interaction term comprised of the effects $\beta_1\beta_2+\tau_j\beta_2\gamma_1$ and the significance of this term was tested using the hypothesis H_{01}: $\beta_1\beta_2=0$ and H_{02}: $\beta_2\gamma_1=0$. The contrast matrices used to test these hypotheses were:

$$C_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (5.23)$$

$$C_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (5.24)$$

The interaction term was tested for significance using the Wald test statistic (see Appendix II (A16) and (A17)). Both tests were based on 1-df, since each contrast matrix has one linearly independent row.

(d) Fitting the unadjusted 'smoke' covariate

The unadjusted 'smoke' covariate could not be fitted, as the constraints were undeterminable. Thus the Asymptotic Relative Precision (APR) could not be computed for this model.

(v) Comparison of the different Cumulative Logit Models

In the literature tests used to compare different cumulative logit models (e.g. to compare the constrained partial proportional odds model with the unconstrained partial proportional odds model or to compare the proportional odds model with the unconstrained partial proportional odds model) are based on the change in deviance. However, for the cumulative models fitted here since weighted least squares were used, an alternative method had to be considered (i.e. the Wald test statistic). In the unconstrained partial proportional odds model, for a given parameter where different log odds were fitted over the cut-points the null hypothesis H_0: $\gamma_{21} = \gamma_{31} = 0$ (since $\gamma_{11} = 0$) was incorporated into the contrast statement, and this assessed whether the proportional odds model was as good a fit as the unconstrained partial proportional odds model. The test was based on 2-df and the C matrix took on the form:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (5.25)$$

Likewise, for a covariate where a trend was apparent in the beta parameters and a set of constraints were considered, the test of whether a model using the constraints was as good a fit as a model using the individually estimated parameters could also be obtained using the contrast statements. This test was set up in the unconstrained partial proportional odds model and for 'smoke' one assessed the null hypothesis H_0: $\gamma_{21} = \Gamma_2\gamma_1$; $\gamma_{31} = \Gamma_3\gamma_1$. (since $\gamma_{11} = \Gamma_1\gamma_1$

=0). As Γ_2=1 and Γ_3=40, then the test was of the form H_0: $\gamma_{31} = 40\gamma_{21}$. The contrast matrix C took on the form:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -40 & 1 & 0 \end{bmatrix} \qquad (5.26)$$

and the test was distributed chi-squared using1-df.

5.2.3.3 Adjacent Category Models

The model fitting procedure for the adjacent category models was similar to that of the cumulative logit models. The first step was to obtain the observed marginal probabilities/response functions and their variance covariance matrix.

(a) Observed sample marginal probabilities

These were as specified for the polytomous model.

(b) Observed sample response functions

The observed sample response functions for the adjacent category model in terms of the health status categories were defined as:

$\ln\left(\dfrac{\Pr(Y_i = y_2)}{\Pr(Y_i = y_1)}\right)$ and this is the log odds based on the '*good*' v. '*excellent*' categories and can be denoted by $f_{i1}(p)$;

$\ln\left(\dfrac{\Pr(Y_i = y_3)}{\Pr(Y_i = y_2)}\right)$ is the log odds based on the '*fair*' v. '*good*' categories and can be denoted by $f_{i2}(p)$;

$\ln\left(\dfrac{\Pr(Y_i = y_4)}{\Pr(Y_i = y_3)}\right)$ is the log odds based on the '*poor*' v. '*fair*' categories and can be denoted by $f_{i3}(p)$.

The observed sample response functions were computed as:

$$f_{i1}(p) = \ln(p_{i2} / p_{i1}); \qquad f_{i2}(p) = \ln(p_{i3} / p_{i2}); \qquad f_{i3}(p) = \ln(p_{i4} / p_{i3});$$

and were derived using F(*p*) =*K* ln(*Ap*), where *A* and *K* were appropriate matrix containing 0, -1 and 1 and having configurations for the adjacent category logits.

The design matrix, *X* depended on the type of adjacent category model fitted. There were two types of these models: (i) constant and (ii) different slope models.

(i) Constant slope Adjacent Category Model

(a) Model specification

The adjacent category model (3.16) as specified by Ananth et al. (1997) and Agresti (1989) has a single constant slope over all the cut-points and this model took on the same form as the proportional odds model (5.5). The design matrix used to model the adjacent category logits was as provided in (5.6).

(b) Assumption of constant odds

Although no test is available to assess whether a constant slopes assumption is appropriate, the cut-point specific parameters for each covariate were tested for homogeneity using the different slopes adjacent category model and contrast statements. The test was based on 2-df for each covariate and the hypotheses were: H_{01}: $\beta_{11}=\beta_{21}=\beta_{31}$ and H_{02}: $\beta_{12}=\beta_{22}=\beta_{32}$. The contrast matrices were given in (5.9) and (5.10) respectively.

(c) Estimation - parameter estimates and odds ratios

The observed sample marginal probabilities, observed sample adjacent category logits and the design matrix as specified in (5.6), were used to solve the weighted least square equations as given in Appendix II. As a result, the parameter estimates and their standard errors were obtained.

The log odds ratios and their standard errors were computed using the parameter estimates. The odds ratios and their 95% confidence intervals followed.

(d) Interaction term

The main effects for the adjacent category model were assumed to have constant slopes over the cut-points and with this in mind the first-order interaction effects were constructed. As the interaction term was also assumed to have a constant slope the form of design matrix for the interaction model was as given in (5.7). The interaction effect was tested for significance using H_0: $\beta_1\beta_2=0$. This was done by setting up a contrast matrix, C which took on the form [0 0 0 0 0 1]. The Wald test statistic (as given in (A17) in Appendix II) was used to test the significance of the interaction effect and this was based on 1-df.

(ii) Different slopes Adjacent Category Model

(a) Model specification

For the adjacent category model (3.17) as defined by Manor et al. (2000), different slope parameters were required over the cut-points. For the Health Status data this model took on a similar form as the polytomous model (5.1) in that the systematic components of both these models were the same. The design matrix for this model was as specified in (5.2).

(b) Estimation - parameter estimates and odds ratios

The weighted least square equations (as defined in Appendix II) were solved using the observed sample marginal probabilities, observed sample adjacent category logits and the design matrix specified in (5.2). The solution resulted in the parameter estimates and their standard errors.

The log odds ratios were obtained in a similar way to that of the polytomous model. Subsequently, the odds ratios and their 95% confidence intervals were computed.

(c) Interaction term

The design matrix used to obtain the parameter estimates of the interaction model was as specified in (5.3). The interaction effects were tested for significance using (A16) in Appendix II with H_0: $\beta_{11}\beta_{12}=0$; $\beta_{21}\beta_{22}=0$; $\beta_{31}\beta_{32}=0$. This test was based on 3-df and was set up in a contrast matrix. The Wald test as specified in (A17) in Appendix II was used to assess the statistical significance.

5.2.3.4 Continuation Ratio Models

From the literature the most effective method that exists to compute the continuation ratio model is that cited by Cole and Ananth (2001). This method fits both the fully unconstrained and unconstrained continuation ratio models. It permits the individual cut-point specific continuation ratio logits to be computed in one programming step and also has the advantage that some of the cut-point specific parameters can be constrained to be the same. The drawback of this method is that one has to transform the data into person-threshold format. Also dummy variables are created in the datafile.

In this thesis, the method described by Scott et al. (1997), whereby separate logistic regression models were computed, was used to fit the different slope continuation ratio model. Also in addition to this, the latter was fitted using the method described for the polytomous model (as cited by Stokes et al., 1985). However, the response functions had to be computed from first principles (as no code in *SAS* exists for doing this). The model was fitted using these and the various design matrices. The resultant was the different versions of the continuation ratio model. This has not been cited anywhere in the literature and is similar to the method suggested by Cole and Ananth (2001). As different versions of the model could be obtained (*unconstrained* – different slopes, *fully constrained* – constant slopes and *partially constrained* – where some of the cut-point specific parameters are constrained to be the same), with little programming efforts.

The observed sample marginal probabilities and response functions with their variance covariance matrices were derived in the usual way. In terms of the observed response categories the continuation ratio logits are:

$\ln\left(\frac{\Pr(Y_i = y_1)}{\Pr(Y_i > y_1)}\right)$ is the log odds based on the '*good*' v. '*excellent*' categories and can be denoted by $f_{i1}(p)$;

$\ln\left(\frac{\Pr(Y_i = y_2)}{\Pr(Y_i > y_2)}\right)$ is the log odds based on the '*fair*' v. ('*excellent*', '*good*') categories and can be denoted by $f_{i2}(p)$;

$\ln\left(\frac{\Pr(Y_i = y_3)}{\Pr(Y_i > y_3)}\right)$ is the log odds based on the '*poor*' v. ('*excellent*', '*good*', '*fair*') categories and can be denoted by $f_{i3}(p)$.

(i) Fully constrained Continuation Ratio Model

Initially the constant slope assumption was checked using the global score test. A significant score test statistic indicated lack of homogeneity. Then each covariate was fitted to check where the heterogeneity existed using the method described in section 5.2.3.2 (ii) d.

(ii) Different slopes Continuation Ratio Model

The different slope continuation ratio model was fitted using two different methods: (i) using separate binary logistic models and (ii) fitting the unconstrained continuation ratio model (similar to that cited by Cole and Ananth (2001)), using the observed probabilities and response functions and the specified design matrix.

(a) Model specification – fitting separate binary logistic models

The results from the separate fits were eventually amalgamated and summarised for the continuation ratio model and therefore there was a need to allow for multiple testing (against the Type I error rate α=0.05/3=0.02; where 3 reflects the three models fitted). In the literature, multiple testing does not feature when carrying out the binary analysis, in relation to the continuation ratio model. However, it was considered relevant here, as results from each model explained some of the overall results for the continuation ratio logits.

(b) Interaction terms– fitting separate binary logistic models

The first order interaction terms were fitted and tested for each of the binary logistic regression models (the test was the Wald statistic based on the χ^2 -distribution with 1-df – although the deviance statistic could also have been used). The corrected Type I error rate was used when testing for significant interaction terms.

(c) Estimation -parameter estimates and the odds ratios – fitting separate binary logistic models

The parameter estimates were obtained using the method of maximum likelihood. The results using the odds ratios and the 95% confidence intervals were amalgamated from the binary analyses to summarise the continuation ratio model.

(d) Model specification – fitting unconstrained continuation ratio model

The design matrix for this model was as for polytomous model given in (5.2).

The specification of the link function usually indicates the type of logits required and in *SAS* there already exist commands that one can use if the logits are based on cumulative odds, adjacent category odds or the polytomous odds models. However no *SAS* code exists for the specification of the continuation ratio logits. To specify the latter logits we refer to the definition of the response functions, $F(\pi) = K\ln(A\pi)$. For the Health Status score, the response is recoded in reverse order (i.e. '*poor*' (1), '*fair*' (2), '*good*' (3) and '*excellent*' (4)) so as that the logits can be computed correctly. The continuation response functions can be expressed as:

$f_{gh1}(\pi) = ln\ (\pi_3/\pi_4)$;

$f_{gh2}(\pi) = ln\ (\pi_2/(\pi_3+\pi_4))$;

$f_{gh3}(\pi) = ln\ (\pi_1/(\pi_2+\pi_3+\pi_4))$.

Here π_1, π_2, π_3 and π_4 are the proportions based on the marginal totals in cells '*poor*', '*fair*', '*good*' and '*excellent*' respectively.

These expressions can be written as:

$f_{gh1}(\pi) = ln(\pi_3) - ln(\pi_4)$;

$f_{gh2}(\pi) = ln(\pi_2) - ln(\pi_4+\pi_3)$;

$f_{gh3}(\pi) = ln(\pi_1) - ln(\pi_4+ \pi_3+\pi_2)$.

In matrix notation, the logits can be expressed as:

$$\begin{bmatrix} 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \ln \pi_1 \\ \ln \pi_2 \\ \ln \pi_3 \\ \ln \pi_4 \\ \ln(\pi_3 + \pi_4) \\ \ln(\pi_2 + \pi_3 + \pi_4) \end{bmatrix}$$

and can be expressed in the form of $F(\pi) = K \ln(A\pi)$ as:

$$\begin{bmatrix} 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \ln \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix}. \qquad (5.27)$$

These matrices can be specified in the RESPONSE statement in *SAS* and allow the computation of the continuation ratio logits.

(e) Estimation – parameter estimates and the odds ratios – fitting unconstrained continuation ratio model

The parameter estimates were obtained by the method of weighted least squares.

The predicted response functions, $F(\hat{\pi})$, were derived and consequently the odds ratios and their 95% confidence intervals resulted using the contrast statement given in section 5.2.3.1 (d).

(f) Interaction term – fitting the unconstrained continuation ratio model

The first order interaction model was fitted using the design matrix specification given in (5.3) and the response functions as specified in (5.27). The test of the interaction term was similar to that described in section 5.2.3.3 (ii) c.

(iii) Partially constrained Continuation Ratio Model

The output from the unconstrained continuation ratio model indicated that some of the cut-point parameters could be constrained to be the same. In testing for homogeneity, there was evidence that for 'smoke' the regression coefficients for cut-points 1 and 2 could be taken as being homogenous. Similarly there was evidence that the regression coefficients for cut-points 2 and 3 of 'heart attack' could be assumed to be homogenous. A partially constrained continuation model was fitted and the formation of the odds was such that $\beta_{11}=\beta_{21}$ and $\beta_{22}=\beta_{32}$. The design matrix was of the form:

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 & -1 \end{bmatrix}.$$

This matrix was used together with the above (5.27) to allow to fit the partially constrained continuation ratio logits.

5.2.3.5 Non-linear Stereotype Model (with unknown constraints)

(a) Model specification

The logits based on the cut-points used for the stereotype model were same as those for the polytomous model. The stereotype model was fitted in *SAS* using Hendrickx's macros (2000) and the detail of how this was done using conditional logistic regression are given in the Appendix II (Section 4). Note that using the macros the $\{\beta_k\phi_j$ s$\}$are not the regression coefficients which are used to provide the log odds but they are the log odds ratios themselves.

The predicted logits of the stereotype model (3.23) for the Health Status data took on the form:

$$F(\pi) = \alpha_j + \sum_{k=1}^{2} x_{ik}\phi_j\beta_k \qquad j=1, 2, 3. \qquad (5.28)$$

This model estimated a scaling metric for the response based on the effects of 'smoke' and 'heart attack'. The model had 3 standard multinomial intercept parameters; it estimated 2 (i.e. (c-2) since $\phi_1 = 0$ and $\phi_4 = 1$) independent ϕ_j scale parameters for the response and a single scaled beta parameter for each independent variable (note that the sign of ϕ_j is positive due to its parameterisation).

<u>*Dimensionality*</u>

The above stereotype model was one-dimensional, as given the covariates x_{i1} and x_{i2}, the same combination of variables $\phi_j\left[\beta_1 x_{i1} + \beta_2 x_{i2}\right]$ could be used to distinguish between all the levels of the outcome. If however, one combination could distinguish between '*excellent*' and '*good*', but a different one was required to distinguish levels '*good*' and '*fair*', the relationship would be described as two-dimensional. To consider the two-dimensional stereotype model, there had to be evidence that one set of predictors were strongly related to some of the response categories, whereas another set of predictors were strongly related to a different part of the response scale. Cross-tabulating 'health status' with the 'smoke' covariate, the chance of having some form of good health status was found to be lower for smokers compared to non-smokers and there was greater chance that those who smoked would have '*poor*' health. Similarly, respondents with heart attack were less likely to have '*excellent*' health and those who had never had a heart attack were less likely to have '*poor*' health. The both covariates appeared to behave in a similar fashion with regards to all the response categories and this was further confirmed by the odds ratios. Therefore the need to look at a 2-dimensional model did not arise.

Indistinguishability

Two health status outcome categories were considered indistinguishable with respect to the covariates, if these covariates were not predictive between the two categories. There was indication that ϕ_3 based on the cut-point '*fair*' v. '*excellent*' was similar to ϕ_4 based on the categories '*poor*' v. '*excellent*' and therefore indistinguishability could be considered. However, since the regression parameters for the final cut-point are constrained to be 1, the parameter based on the cut-point ('*fair*' v. '*excellent*') would also have to be constrained to be 1 in order to consider indistinguishability. Although Hendrickx's macros (2000) fit the stereotype model by estimating the scaling and regression parameters, there is no in-built mechanism to test whether any adjacent categories are similar and if so, whether these can be constrained to take on similar parameter values. However, Lunt (2001) has devised macros (in *SAS* and *Stata*) known as *Soreg*, which allow the fit of more than one-dimension of the stereotype model and also consider aspects of indistinguishability (see 3.5.6 (e)). Although the latter macro fits the stereotype model allowing for indistinguishability, one cannot compare the two models (i.e. with and without indistinguishability). Unlike for generalised linear models, it cannot be shown that the log-likelihoods follow a χ^2 distribution asymptotically for the former models. The deviance statistic (i.e. change in the –2log-likelihoods of the two models) cannot be used to compare the two stereotype models with the different constraints. The comparison of the two models, in theory could be done using, bootstrapping technique (see below). However, Lunt's macros were not compatible with the bootstrap techniques presented in *Stata*. Therefore the indistinguishability of the models could not be tested.

(b) Estimation- parameter estimates and odds ratios

Due to the non-linear nature of the stereotype models and therefore the estimation of the $\{\beta_k\phi_j$ s$\}$ parameters, the standard errors of the parameter estimates $\{\beta_k\}$ were conditional given the scaling metric ϕ_j and therefore these were not valid. Likewise any inference based on the standard errors or the likelihood-based tests was also not correct. To obtain valid standard errors for the parameters and therefore for the log odds and odds ratios, bootstrapping was used.

The bootstrap technique (Efron and Tibshirani, 1993) involved repeated re-estimation of a parameter using random samples with replacement from the original dataset. In the case of the Health Status data, 100 bootstrap samples were drawn and each one was used to fit the stereotype model. For each model the parameters β_k and ϕ_j were estimated. Five parameter ($\phi_2, \phi_3, \phi_4, \beta_1, \beta_2$) distributions were obtained. Each distribution could be assumed to be normal and therefore the mean and standard error were calculated to give the point estimates of the β_k and ϕ_j together with their standard errors. The log odds ratios $\beta_k\phi_j$ were also obtained from the parameters and again there were 100 values for each log odds $\beta_k\phi_j$; k=2, 3, 4; j=1, 2 (since $\phi_1 = 0$). Each log odds took on a distribution form from which the mean and standard error were determined. These values provided the point estimate of the log odds ratio and its standard error.

(c) Interaction term

To assess whether the 1st order interaction term was significant or not, the usual change in the –2log-likelihood between the main effects and saturated models could not be carried out due to the conditional parameters. Bootstrapping was used instead. The saturated and the main effect models were fitted using each bootstrap sample and the change in the –2log-likelihoods was obtained. As there were a 100 of these samples, there were 100 change values that formed a distribution. The observed change value was compared in this distribution. The null hypothesis was that there was no difference in the two models (i.e. the interaction term was not significant). The ASL (Achieved Significance Level) for the bootstrap test was the proportion of the number of change values in the distribution that was greater than or equal to the observed change value, i.e. #(change value in distribution ≥ observed change value)/100, where 100 was the number of bootstrap samples. A small ASL implied that the null hypothesis was rejected and that the interaction model was the preferred model.

(d) The APR of the 'smoke' covariate

Due to the non-linear nature of the stereotype model, the calculation of the ARP was not applicable and therefore this statistic was not obtainable for this model.

5.2.3.6 The Linear Stereotype Model (with known constraints)

The constraints $\{\phi_j\}$ were obtained as estimated parameters using Hendrickx (2000) macros. These constraints were taken as constants and the linear form of the stereotype model with β regression parameters for each covariate were fitted. The purpose of this latter model was related to model checking. The goodness-of-fit of the stereotype model which had been fitted using Hendrickx's (2000) or Lunt's (2001) macros could not be easily checked due to the non-linear nature of the model. By using the estimated constraints, the stereotype model was of a linear form and the usual estimated response functions and probabilities could be assessed.

(a) Observed sample marginal probabilities and response functions

The observed sample marginal probabilities and response functions were as for the polytomous model (specified in section 5.2.3.1).

(b) Model specification

The design matrix for the linear version of model (5.28) is as in the matrix formulation:

$$\begin{bmatrix} f_{111}(\hat{\pi}) \\ f_{112}(\hat{\pi}) \\ f_{113}(\hat{\pi}) \\ f_{121}(\hat{\pi}) \\ f_{122}(\hat{\pi}) \\ f_{123}(\hat{\pi}) \\ f_{211}(\hat{\pi}) \\ f_{212}(\hat{\pi}) \\ f_{213}(\hat{\pi}) \\ f_{221}(\hat{\pi}) \\ f_{222}(\hat{\pi}) \\ f_{223}(\hat{\pi}) \end{bmatrix} = \begin{bmatrix} \alpha_1 + c_1\beta_1 + c_1\beta_2 \\ \alpha_2 + c_2\beta_1 + c_2\beta_2 \\ \alpha_3 + c_3\beta_1 + c_3\beta_2 \\ \alpha_1 + c_1\beta_1 - c_1\beta_2 \\ \alpha_2 + c_2\beta_1 - c_2\beta_2 \\ \alpha_3 + c_3\beta_1 - c_3\beta_2 \\ \alpha_1 - c_1\beta_1 + c_1\beta_2 \\ \alpha_2 - c_2\beta_1 + c_2\beta_2 \\ \alpha_3 - c_3\beta_1 + c_3\beta_2 \\ \alpha_1 - c_1\beta_1 - c_1\beta_2 \\ \alpha_2 - c_2\beta_1 - c_2\beta_2 \\ \alpha_3 - c_3\beta_1 - c_3\beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & c_1 & c_1 \\ 0 & 1 & 0 & c_2 & c_2 \\ 0 & 0 & 1 & c_3 & c_3 \\ 1 & 0 & 0 & c_1 & -c_1 \\ 0 & 1 & 0 & c_2 & -c_2 \\ 0 & 0 & 1 & c_3 & -c_3 \\ 1 & 0 & 0 & -c_1 & c_1 \\ 0 & 1 & 0 & -c_2 & c_2 \\ 0 & 0 & 1 & -c_3 & c_3 \\ 1 & 0 & 0 & -c_1 & -c_1 \\ 0 & 1 & 0 & -c_2 & -c_2 \\ 0 & 0 & 1 & -c_3 & -c_3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad (5.29)$$

In (5.29) the c_j are the scalars from the estimated parameters ϕ_js. The predicted logits can be expressed as specified in Table 5.5.

In this model five parameters were fitted and their definition was similar to that of the proportional odds model.

(c) Estimation –parameter estimates and odds ratios

The parameter estimates were obtained by maximum likelihood estimation (as detailed in Appendix II (A11)). These estimates (and their standard errors) were compared to those obtained from the non-linear stereotype model fitted using Hendrickx's (2000) macros and the bootstrap technique. The predicted response functions $F(\hat{\pi})$ were obtained by solving (5.29) and these together with the predicted probabilities were used to assess the goodness-of-fit of the stereotype model.

Table 5.5: Table of model predicted log odds for the Stereotype Model with known constraints using the Health Status data

Do you smoke? (g)	***Have you had a heart attack? (h)***	$f_{gh1}(\hat{\pi})$	$f_{gh2}(\hat{\pi})$	$f_{gh3}(\hat{\pi})$
Yes	***Yes***	$\alpha_1 + c_1\beta_1 + c_1\beta_2$	$\alpha_2 + c_2\beta_1 + c_2\beta_2$	$\alpha_3 + c_3\beta_1 + c_3\beta_2$
Yes	***No***	$\alpha_1 + c_1\beta_1 - c_1\beta_2$	$\alpha_2 + c_2\beta_1 - c_2\beta_2$	$\alpha_3 + c_3\beta_1 - c_3\beta_2$
No	***Yes***	$\alpha_1 . c_1\beta_1 + c_1\beta_2$	$\alpha_2 - c_2\beta_1 + c_2\beta_2$	$\alpha_3 - c_3\beta_1 + c_3\beta_2$
No	***No***	$\alpha_1 - c_1\beta_1 - c_1\beta_2$	$\alpha_2 - c_2\beta_1 - c_2\beta_2$	$\alpha_3 - c_3\beta_1 - c_3\beta_2$

(d) The Asymptotic Relative Precision of the 'smoke' covariate

The Asymptotic Relative Precision (ARP) was not computed for the same reasons stated for the non-linear stereotype model.

5.3 Townsend Disability Scale Data

The covariates 'sex' and 'full-time education' were taken to be categorical and 'age group' was taken to be ordinal (except in the binary model). Interest was focused on the 'full-time education' covariate and therefore the other covariates were adjusted for this latter one. In the models fitted using this dataset, first order interaction terms were considered. Three interaction terms were assessed individually – 'age group *x* sex', 'sex *x* full-time education' and 'age group *x* full-time education'. The model strategy was different using this dataset compared to that adopted for the Health Status data. The global χ^2-score test statistic was of little value as interest was focused on 'full-time education' and it was important that proportionality or constant odds be primarily satisfied by this covariate (as opposed to all the covariates). Thus this test was not used for testing constant odds.

5.3.1 Linear Regression Model

The assumptions stated in section 5.2.1 were considered for this model. The analysis using linear regression was based on complete observations for all covariates and response and 12434 observations were available. Forward selection was used and initially each covariate was fitted individually into the model. The 1st order interaction terms were tested with the main effects using the *t-test statistic* (H_0: parameter = 0) each of which was t-distributed with 1 df. The parameter estimates for the final model were obtained.

5.3.2 Binary Logistic Regression Model

McGee et al. (MRC CFAS[1], 1998) took the total disability score (summed over the 9 items) for a given participant and computed it as a proportion of the total score (which was equal to 18). This proportion was then treated in the logistic regression model as the dependent variable. The nature of this response variable indicated *over-dispersion*, where the data did not fit the binomial distribution very well mainly as a result of too much random variation (assessed by the goodness-of-fit statistic, i.e. deviance/df). Over-dispersion was corrected for by adjusting/scaling the covariance/variance matrix. This involved the estimation of an additional parameter, the scaling parameter. By adjusting for over-dispersion, the mean and variance of the score were correctly fitted.

The analyses carried out by McGee et al. (MRC CFAS[1], 1998) were based on a single covariate model and a multivariate model containing nine covariates, each with 'sex' and 'age group' (taken as categorical) as adjusting variables.

There were three reasons why these analyses were replicated: (i) the data used in the publication were an older version (with 12114 observations) of the one used in this thesis and therefore more observations were available; (ii) nine covariates were used in the publication. For the models fitted here only three covariates were chosen namely 'sex', 'age group' and 'full-time education' (on two levels as opposed to four as presented in the publication). Therefore there was a need to re-examine the results using only three covariates; (iii) the interpretation of the results provided in the publication was not very clear.

The binary logistic model for the Townsend Disability Score data in this thesis was fitted using McGee et al.'s (MRC CFAS[1], 1998) method. Thus the models here were all fitted with a scaling parameter to correct for over-dispersion. The main effects model was fitted and then each interaction term was added and tested for significance. For this analysis, 'age group' was taken as categorical (see Appendix II: section 1) and had 5 categories. The 'age group *x* sex ' term was tested using 4-df; 'sex *x* full-time education' was tested using 1-df and the test for 'age group *x* full-time education' had 4-df. The maximum likelihood procedure was used with 'sex' and 'age group' as adjusting covariates and interest was focused around the 'full-time education' covariate. The adjusted odds ratios for the 'full-time education' together with the 95% confidence interval of the odds ratio were obtained.

5.3.3 Ordinal Regression Models

5.3.3.1 Polytomous Model

As for the Health Status data, the observed marginal probabilities and response functions were computed initially. The design matrix required for modelling, depended on the model specification.

(a) Model specification

If we assume that an individual in group i takes on the values for 'full-time education' f= 1(<13 years) and 2=($\geq$ 13 years); 'sex' g=1 (males) 2 (females) and 'age-group' h=1 (< 70), 2 (70-75), 3(75-80), 4 (80-85) and 5($\geq$ 85), then given the referent category is '*severe* + *v. severe*' and i=fgh sub-population, we can write the predicted generalised logits (using (3.6)) for the Townsend disability score data as:

$\ln\left(\frac{\Pr(Y_i = y_1)}{\Pr(Y_i = y_5)}\right)$ is the log odds for categories '*none*' v. '*severe + v. severe*' and can be denoted by $f_{i1}(\hat{\pi})$;

$\ln\left(\frac{\Pr(Y_i = y_2)}{\Pr(Y_i = y_5)}\right)$ is the log odds for categories '*slight*' v. '*severe + v. severe*' and can be denoted by $f_{i2}(\hat{\pi})$;

$\ln\left(\frac{\Pr(Y_i = y_3)}{\Pr(Y_i = y_5)}\right)$ is the log odds for categories '*some*' v. '*severe + v. severe*' and can be denoted by $f_{i3}(\hat{\pi})$;

$\ln\left(\frac{\Pr(Y_i = y_4)}{\Pr(Y_i = y_5)}\right)$ is the log odds for categories '*appreciable*' v. '*severe + v. severe*' and can be denoted by $f_{i4}(\hat{o})$.

The polytomous model took on a similar form to (5.1). For this model sixteen parameters were fitted: four $\{\alpha_j\}$ parameters were the nuisance parameters; four β_{j1} parameters were the differential change in the log odds for 'age group' (with the linear constraint imposed on the levels of 'age group') for each of the cut-points; four β_{j2} cut-point specific parameters were the differential change in the log odds for 'sex' (males/females) and four β_{j3} were the cut-point specific parameters which were the added amount for less than 13 years of full-time education and the subtracted amount for more than 13 years of full-time education.

The design matrix for this model was obtained in a similar fashion to that of the Health Status data, and can be derived from the log odds table (Table IIa) in Appendix II.

(b) Estimation – parameter estimates and odds ratios

The parameter estimates were obtained by the method of maximum likelihood. This method involved computing a function of the observed marginal probabilities, response functions and the design matrix. The predicted response functions were obtained and consequently the odds ratios and their 95% confidence interval.

(c) Interaction terms

The 1st order interaction terms were constructed using the product of the main effects. Three interaction terms were assessed individually. The effects of each interaction term were incorporated into the design matrix of the main effects. The inclusion of each term was assessed using the null hypothesis that all the interaction effects were statistically significant from zero and using the Wald test statistic (and this test was based on 4-df). The contrast statement was used to perform the tests for the interaction terms. (Note: the test based on the change in the –2log-likelihood between the main effects model and that with an interaction term (i.e. the deviance) could have been used as an alternative as it is an equally efficient way of assessing significance of the interaction term).

5.3.3.2 Cumulative Logit Models

The observed marginal probabilities and response functions were obtained in a similar way to that given in section 5.2.3.2. The observed sample logits were denoted as:

$\ln\left(\frac{\Pr(Y_i > y_1)}{\Pr(Y_i \le y_1)}\right)$ was the cumulative logit for the response categories '*none*' v. ('*slight*', '*some*', '*appreciable*', '*severe + v. severe*') and was denoted as $f_{i1}(p)$;

$\ln\left(\frac{\Pr(Y_i > y_2)}{\Pr(Y_i \le y_2)}\right)$ was the cumulative logit for the response categories ('*none*', '*slight*') v. '*some*', '*appreciable*', '*severe + v. severe*') and was denoted as $f_{i2}(p)$;

$\ln\left(\frac{\Pr(Y_i > y_3)}{\Pr(Y_i \le y_3)}\right)$ was the cumulative logit for the response categories ('*none*', '*slight*', '*some*') v.('*appreciable*', '*severe + v. severe*') and was denoted as $f_{i3}(p)$ and

$\ln\left(\frac{\Pr(Y_i > y_4)}{\Pr(Y_i \le y_4)}\right)$ was the cumulative logit for the response categories ('*none*', '*slight*', '*some*', '*appreciable*') v. '*severe + v. severe*' and was denoted as $f_{i4}(p)$.

The formulation of the design matrix depended on the type of cumulative logit model fitted.

(i) Different slopes Cumulative Logit Model

The different slopes model fitted for the Townsend Disability Score data was different to that fitted using the Health Status data due to the different modelling strategy chosen and due to the fact that attention was focused on 'full-time education'.

(a) Model specification

The purpose of this model was to assess the behaviour of the parameter estimates over the cut-points and also to decide whether proportional odds existed for full-time education. Therefore initially each covariate was tested individually for proportional odds by using the hypothesis based on the cut-points - H_0: $\beta_{1k}=\beta_{2k}=\beta_{3k}=\beta_{4k}$ (k= 'sex', 'age group' and 'full-time education'). The contrast statement was used to construct and test the hypotheses and each test was based on 3-df (using (A16) and (A17) in Appendix II). The contrast matrix, C, was of the form:

$$C=\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix}. \qquad (5.30)$$

(b) Model assumption

If we fitted the different slopes model as for the Health Status data, then the estimates of full-time education would assume different slopes for the adjusting covariates. Thus, to include the covariates in the model, forward selection was used. The adjusting covariates 'age group' and 'sex' had to be tested for proportional odds prior to fitting in the 'full-time education' variable. The covariates that had proportional odds were kept in the model. It was found that whilst proportional odds existed for 'age group' and 'sex', non-proportionality was present for 'full-time education'.

(c) Estimation - parameter estimates and odds ratios

The parameter estimates for this model were obtained using weighted least square equation (as in Appendix II). The odds ratios with their 95% confidence intervals were *not* obtained for this model (as the purpose here was to assess how the covariates were behaving over the cut-points).

(d) Interaction terms

The different slopes model was primarily used to assess the behaviour of the main effects and therefore the interaction terms were not considered relevant.

(ii) Proportional Odds Model

Although the proportional odds assumption was not satisfied, the model was fitted, so as that the estimates of the parameters could be compared with other models. Each interaction term was assumed to have proportional odds and fitted and tested for significance (test based on 1 df). The appropriate design matrix for the final model was formed and used in the weighted least square equations to obtain the parameter estimates.

(iii) Unconstrained Partial Proportional Odds

(a) Model specification

The 'full-time education' covariate demonstrated some evidence of non-proportionality and therefore a partial proportional odds model was more feasible, than a proportional odds model. The main effects 'sex' and 'age group' were fitted with proportional odds whilst 'full-time education' was fitted with non-proportional odds. The model took on the form (given group i has fgh covariate characteristics or falls in that sub-population):

$$F(\pi) = \alpha_j + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + T_{i3}\gamma_{j3}. \qquad (5.31)$$

Each first-order interaction term was added to this main effects model, such that the models fitted were:

$$F(\pi) = \alpha_j + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + T_{i3}\gamma_{j3} + x_{i4}\beta_4; \qquad (5.32)$$

$$F(\pi) = \alpha_j + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + T_{i3}\gamma_{j3} + x_{i4}\beta_4 + x_{i5}\beta_5 + T_{i5}\gamma_{j5}; \qquad (5.33)$$

$$F(\pi) = \alpha_j + x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + T_{i3}\gamma_{j3} + x_{i4}\beta_4 + x_{i6}\beta_6 + T_{i6}\gamma_{j6}. \qquad (5.34)$$

Here x_{i4} represents the interaction 'age group *x* sex'; x_{i5} and Γ_{i5} represents the interaction 'age group *x* full-time education' and x_{i6} and Γ_{i6} represents the interaction 'sex *x* full-time education'. The interaction 'age group *x* sex' was made up of one effect (β_4) as the proportional odds were assumed for the main effects. The interaction terms of 'sex' and 'full-time education' and 'age-group' and 'full-time education' did not assume proportional odds over the cut-points and therefore the regression coefficients for these were made up of two components: the constant slopes component and the left over component which varied by cut-point.

The hypotheses H_{01}: β_4=0 (for 'age group *x* sex') was set up to test the significance of the effect. Also H_{02}: β_5=0 and H_{03}:$\gamma_{25}=\gamma_{35}=\gamma_{45}=0$ (since $\gamma_{15}=0$) was constructed for 'age group *x* full-time education' and the hypotheses H_{04}: β_6=0 and H_{05}:$\gamma_{26}=\gamma_{36}=\gamma_{46}=0$ (since $\gamma_{16}=0$) were set up for 'sex *x* full-time education'. Note that for the above models γ_{13} was also set to zero. Initially the interaction 'age group *x* sex' was found to be significant. Then the interaction of 'sex *x* full-time education' was found significant, allowing for the terms in model (5.32). The final model could be represented as (5.34).

In model (5.34) 15 parameters were fitted (two were set to zero): four$\{\alpha_j\}$ parameters, one constant slope parameter for 'age group', one constant slope parameter for 'sex', one constant slope parameter for 'full-time education' with four cut-point specific parameters which accounted for the non-proportionality (with γ_{13}=0). There was also one interaction parameter for 'age group *x* sex', and five interaction parameters for 'sex *x* full-time education' (where one of these parameters was the constant component of the interaction and the other four were the ones that varied by cut-point and γ_{16}=0).

The predicted log odds formation for this model took on the form as specified in Appendix II (Table IIc). From this the design matrix could be derived to fit the model.

(b) Estimation - parameter estimates and odds ratios

The parameter estimates were obtained by solving the weighted least squares equations (as in Appendix II). Consequently the odds ratios and their 95% confidence intervals were derived.

(iv) Constrained Partial Proportional Odds Model

The constrained partial proportional odds model was not considered here, as a simple suitable set of constrained that satisfied both the 'full-time education' main effect and the interaction term of 'sex *x* full-time education' could not be found.

5.3.3.3 Adjacent Category Models

The observed sample adjacent category logits were expressed as:

$\ln\left(\dfrac{\Pr(Y_i = y_1)}{\Pr(Y_i = y_2)}\right)$ and this was the adjacent logit for the response categories '*none*' v. '*slight*'

and was denoted as $f_{i1}(p)$;

$\ln\left(\dfrac{\Pr(Y_i = y_2)}{\Pr(Y_i = y_3)}\right)$, this was the adjacent logit for the response categories '*slight*' v. '*some*'

and was denoted as $f_{i2}(p)$;

$\ln\left(\dfrac{\Pr(Y_i = y_3)}{\Pr(Y_i = y_4)}\right)$, this was the adjacent logit for the response categories '*some*' v.

'*appreciable*' and was denoted as $f_{i3}(p)$;

$\ln\left(\dfrac{\Pr(Y_i = y_4)}{\Pr(Y_i = y_5)}\right)$, this was the adjacent logit for the response categories '*appreciable*' v.

'*severe + v. severe*' and was denoted as $f_{i4}(p)$.

The design matrix depended on the type of adjacent category logits fitted.

(i) Constant slope Adjacent Category Model

(a) Model specification

The assumption of a constant slope for each covariate over the cut-points was checked when fitting the constant slope adjacent category model. The test of constant slopes was similar to that described in section 5.3.3.2 (i) (a) and was based on 3-df for each term (fitting the different slope model). There was a requirement for the adjusting covariates 'age group' and 'sex' to have constant slopes, whilst adding in 'full-time education'. This was not found to be the case, and it was found that homogeneity for the cut-point specific parameters existed for 'full-time education', allowing for the different slopes for the adjusting covariates. As a result the adjusting covariates were 'forced' to have constant slopes and 'full-time education' continued to satisfy the model assumption. Therefore, all the covariates were assumed to have constant odds and the required design matrix was used (as for the proportional odds model).

<u>(b) Estimation - parameter estimates and odds ratios</u>

The parameter estimates were obtained using weighted least squares method and using these the odds ratios and their 95% confidence intervals were derived.

<u>(c) Interaction terms</u>

For the main effects model, the first-order interaction terms were constructed and then incorporated into the design matrix. The assumption of constant slopes in the main effects was carried through into the interaction terms. Each interaction term was tested for significance (H_0: Interaction =0) using a contrast statement (and the test was based on 1-df) and the Wald test statistic.

(ii) Different slopes Adjacent Category Model

<u>(a) Model specification</u>

Different parameter estimates are required over the cut-points for each of the covariates for the different slopes adjacent category model. The form of the design matrix was based on the logits as specified in Table IIa (appendix II) and the number of parameters fitted was as for the polytomous model. There were no model assumptions to satisfy, and therefore the model could be fitted without any prior statistical testing. Initially the main effects were fitted. Then each first-order interaction term was added into the model and assessed for significance (H_0: Interaction =0). The test of each interaction term was based on 4-df, since the term was made up of four cut-points.

<u>(b) Estimation - parameter estimates and odds ratios</u>

Weighted least square parameters were obtained. The log odds ratios and their standard errors were derived using the appropriate contrast matrices in (A16) and (A17) in Appendix II.

Interest was focused on the adjusted odds ratios and the 95% confidence intervals for 'full-time education' where less than 13 years education was compared with 13 or more years full-time education.

5.3.3.4 Continuation Ratio Models

The observed and predicted marginal probabilities and response function are derived in a similar way to the above. The observed sample continuation ratio logits were derived as:

$\ln\left(\frac{\Pr(Y_i = y_1)}{\Pr(Y_i > y_1)}\right)$ and this was the continuation ratio logit for the response categories '*none*' v. ('*slight*', '*some*', '*appreciable*', '*severe+ v. severe*') and was denoted as $f_{i1}(p)$;

$\ln\left(\frac{\Pr(Y_i = y_2)}{\Pr(Y_i > y_2)}\right)$ and this was the continuation ratio logit for the response categories '*slight*' v. ('*some*', '*appreciable*', '*severe+ v. severe*') and was denoted as $f_{i2}(p)$;

$\ln\left(\frac{\Pr(Y_i = y_3)}{\Pr(Y_i > y_3)}\right)$ and this was the continuation ratio logit for the response categories '*some*' v. ('*appreciable*', '*severe+ v. severe*') and was denoted as $f_{i3}(p)$;

$\ln\left(\frac{\Pr(Y_i = y_4)}{\Pr(Y_i > y_4)}\right)$ and this was the continuation ratio logit for the response categories '*appreciable*' v. '*severe+ v. severe*' and was denoted as $f_{i4}(p)$.

(i) Fully constrained Continuation Ratio Model

This was fitted in a very similar way to the constant slope adjacent category model.

(ii) Different slope Continuation Ratio Model

Two different methods were used to fit the different slope continuation ratio model (i) the separate binary logistic regression models (ii) the unconstrained continuation ratio model.

(a) Model specification – fitting the separate binary logistic models

Again separate binary logistic regression models were fitted to each cut-point based on the continuation ratio logits. Overall results from these models were summarised using multiple testing. Since four binary logistic regression models were fitted, the Type I error rate was $\alpha=0.05/4=0.01$.

(b) Interaction terms– fitting the separate binary logistic models

Each interaction term (for each binary logistic regression model) was fitted and the corrected Type I error rate was used for multiple testing when testing the significance of each term using the Wald test statistic. Significant interaction terms were included in the binary logistic model (test for each term was based on 1-df).

(c) Parameter estimates and odds ratios– fitting the separate binary logistic models

The parameter estimates from each binary logistic regression model were obtained using the method of maximum likelihood. Again, from these parameter estimates the odds ratios and their 95% confidence intervals were derived. The results based on the odds ratios and their confidence intervals were amalgamated to give an overall result for the continuation ratio model.

(d) Model specification – fitting unconstrained continuation ratio model

The continuation ratio logits for the Townsend Disability Scale can be expressed as:

$$f_{i1}(\pi) = ln(\pi_1/(\pi_2+\pi_3+\pi_4+\pi_5));$$

$$f_{i2}(\pi) = ln(\pi_2/(\pi_3+\pi_4+\pi_5));$$

$$f_{i3}(\pi) = ln(\pi_3/(\pi_4+\pi_5));$$

$$f_{i4}(\pi) = ln(\pi_4/\pi_5),$$

and these can be rewritten as:

$$f_{i1}(\pi) = ln(\pi_1) - ln(\pi_2+\pi_3+\pi_4+\pi_5);$$

$$f_{i2}(\pi) = ln(\pi_2) - ln(\pi_3+\pi_4+\pi_5);$$

$f_{i3}(\pi) = ln(\pi_3) - ln(\pi_4 + \pi_5);$

$f_{i4}(\pi) = ln(\pi_4) - ln(\pi_5).$

In matrix notation one can write:

$$\begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \ln \pi_4 \\ \ln \pi_3 \\ \ln \pi_2 \\ \ln \pi_1 \\ \ln(\pi_2 + \pi_3 + \pi_4 + \pi_5) \\ \ln(\pi_3 + \pi_4 + \pi_5) \\ \ln(\pi_4 + \pi_5) \\ \ln(\pi_5) \end{bmatrix}$$

and in the form $F(\pi) = K \ln(A\pi)$, the logit functions can be written as:

$$\begin{bmatrix} 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \ln \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \end{bmatrix}.$$

Thus this matrix formulation was expressed in the RESPONSE statement in *SAS* and permitted the computation of the continuation ratio logits.

(e) Interaction terms – fitting unconstrained continuation ratio model

Each 1st order interaction term was constructed using the product of the main effects (as for the polytomous model). The test of each interaction term was based on the 4-df. This hypothesis was tested using the Wald test statistic.

(f) Estimation – parameter estimates and the odds ratios – fitting unconstrained continuation ratio model

The parameter estimates and their standard errors were obtained using weighted least squares. The odds ratios and their 95% confidence intervals were derived using these.

(ii) Partially constrained Continuation Ratio Model

The need to fit this model did not arise as it was found that the fully constrained model was appropriate.

5.3.3.5 Non-linear Stereotype Model (with unknown constants)

(a) Model specification

The Townsend disability ordinal score and the three main effects were used to fit the one-dimension stereotype model using *mclest* (Hendrickx, 2000) in *SAS*. The cut-points used for the stereotype model were as for the polytomous model. The 1st order interaction terms were created using the cross product of the main effects. There were three 1st order interaction terms to test: 'age group *x* sex', 'sex *x* full-time education' and 'age group *x* full-time education'. The usual method of comparing each 1st order interaction model with the main effect model could not be used, due to the fact that the –2log-likelihood did not necessarily follow a chi-squared distribution asymptotically. As a result bootstrapping was used. The null hypothesis was based on no differences in the compared models (i.e. the interaction term model was similar to the main effects model). The observed change in the –2log-likelihoods from the main effects model and each interaction term model was obtained. Then a 100 bootstrap samples were used to fit each interaction term and the main effects models. The 100 change values from the –2log-likelihoods of each interaction model and the main effects model formed a distribution. The ASL formed a test of rejecting or accepting the null hypothesis. Thus the proportion #(change value in distribution ≥ observed change value)/100 was computed. In the case of a small ASP, the null hypothesis was rejected and the interaction model was the preferred model.

Dimensionality

The aspect of dimensionality - as for the previous dataset - was examining using the individual unadjusted covariate and the response variable. For a two-dimensional stereotype model one needed to address the question - do different discriminators vary between the response categories (i.e. was there a discriminator for response categories say, 1 and 2 and another discriminator for response categories 2 and 3, 3 and 4 and 4 and 5, say)? From the data, for the 'sex' covariate the chance of '*none*' or '*some*' disability was higher for males than females. Females were more likely to have '*some*', '*appreciable*' or '*severe+ v. severe*' forms of disability. The number of respondents in the '*none*' disability group was higher for those among the younger age groups and as age increased the chance of having a greater form of disability increased too. Respondents with 13 or more years of full-time education were likely to have no disability and those with less education (< 13 years) were more likely to have a greater form of disability (i.e. have '*some*', '*appreciable*' or '*severe+ v. severe*'). There was an indication that all three covariates appeared to behave similarly over the response categories. As a result the need to fit a two-dimensional model did not arise.

Indistinguishability

There was no indication of indistinguishability when assessing the ordering parameters using the one-dimension stereotype model. Therefore this aspect of the stereotype model was not considered further.

(b) Estimation - parameter estimates and the odds ratios

Bootstrapping was carried out on a hundred samples (drawn with replacement) to obtain the corrected standard errors for the parameters and the log odds ratios. Again, bootstrapping was used to obtain the point estimate of log odds ratios, $\phi_j\beta_k$ together with the corresponding standard errors. The odds ratio and their 95% confidence intervals were obtained from these

(c) Comparison of the Stereotype Model with the Polytomous Model for both datasets

The change in the –2log-likelihoods for the polytomous model and the stereotype model provided a way of establishing whether a model with constraints was as good a fit as a model with no constraints imposed on a set of covariates. The null hypothesis was based on no difference in the two models. The observed change in the –2log-likelihood of the polytomous and the stereotype models was computed. Then a 100 bootstrap samples were taken and each sample was used to fit the stereotype and the polytomous model, with the change in the –2log-likelihood being derived using the two models. The 100 change values formed a distribution. The null hypothesis was rejected if the ASL was small.

5.3.3.6 Linear Stereotype Model (with known constraints)

(a) Model specification

The $\{\phi_j\}$ parameters were taken as constants c_j, from the non-linear version of the stereotype model and a linear model was fitted. For this model β_1 was the regression coefficient for 'sex', β_2 was the regression coefficient for 'age group', β_3 was the regression coefficient for 'full-time education' to be estimated. The interaction of 'age group x sex' was found to be significant. Thus β_4 was the regression coefficient for the interaction of 'age group x sex'.

Using this model specification, the design matrix was derived and the model fitted using the method of maximum likelihood. The parameter estimates and the predicted logits/probabilities were obtained and subsequently used for assessing the goodness-of-fit of the stereotype model.

5.4 Issues addressed in this chapter

The issues highlighted at the beginning of this chapter have been addressed. In brief, these include:

(i) *fitting several covariates:* This is straightforward when fitting ordinal regression models with different slopes. Care has to be taken particular for models where the assumptions need to be satisfied.

(ii) *Fitting interaction terms:* For all the ordinal regression models, the interaction terms were constructed using the product of the main effects. For the constant slope model, the assumptions were carried across to the interaction terms.

(iii) *Fitting Partial Proportional Odds Model:* The method based on first principles described in this chapter for fitting ordinal regression models (i.e. using the observed probabilities, response functions and specified design matrix) was used to fit the partial proportional odds model.

(iv) *Unconstrained Continuation Ratio Model:* Again, using the method describe in this chapter, it was relatively straightforward to fit this model. This method did not require re-arranging the data (as for the method given be Cole and Ananth (2001)).

(v) *The Linear Stereotype Model:* The linear version of the stereotype model could be fitted and statistical inference which is not possible on the non-linear version be carried out.

(vi) *Townsend Disability Scale:* An ordinal analysis using the Townsend disability score was carried out. A comparison of the results from this analysis and that provided by the binary logistic regression model (similar to the one in the publication (MRC CFAS[1] (1998))) could be carried out.

Table 5.6: Summary of the Statistical Methods used to fit the Regression Models

Model	Method	SAS/STATA Procedure	Health Status score		5-grouped Townsend Disability Scale categories	
			Cut-points	*Number of Parameters estimated*	*Cut-points*	*Number of Parameters estimated*
Linear Regression	Least squares method	SAS - PROC GLM	Response taken as continuous	3 (with main effect); 4 (with interaction).	Response taken as continuous	4 (with main effects); 5 (with one 1st order inter.).
Binary Logistic Regression	Maximum Likelihood	SAS - PROC LOGISTIC	3 separate models: ('good', 'fair', 'poor') v. 'excellent'; ('fair', 'poor') v. ('excellent', 'good'); 'poor' v. ('excellent', 'good', 'fair').	Each model: 3 (with main effects); 4 (with interaction).	Replication of McGee et al.'s (MRC CFAS[1], 1998) analysis	4 (with main effects); 5 (with one 1st order inter.).
Polytomous	Maximum Likelihood	SAS - PROC CATMOD (logit -response)	'good' v. 'excellent'; 'fair' v. 'excellent'; 'poor' v. 'excellent'.	9 (with main effects); 12 (with interaction).	'none' v. 'severe+v.severe'; 'slight' v. 'severe+v.severe'; 'some' v. 'severe+v.severe'; 'appreciable' v. 'severe+v.severe'	16 (with main effects); 20 (with one 1st order inter.).
Proportional Odds	Weighted least squares	SAS - PROC CATMOD (clogit -response) PROC LOGISTIC (proportional odds assumption)	('good', 'fair', 'poor') v. 'excellent'; ('fair', 'poor') v. ('excellent', 'good'); 'poor' v. ('excellent', 'good', 'fair').	5 (with main effects); 6 (with interaction).	'none' v.('slight', 'some', 'appreciable', 'severe+v.severe'); ('none', 'slight') v. ('some','appreciable','severe+v.severe'); ('none', 'slight','some') v. ('appreciable','severe+v.severe'); ('none','slight','some','appreciable') v. 'severe+v.severe';	7 (main effects); 8 (with interaction)
Unconstrained Partial Proportional Odds	Weighted least squares	SAS - PROC CATMOD (clogit -response)	('good', 'fair', 'poor') v. 'excellent'; ('fair', 'poor') v. ('excellent', 'good'); 'poor' v. ('excellent', 'good', 'fair').	7 (with main effects); 10 (with interaction).	'none' v.('slight', 'some', 'appreciable', 'severe+v.severe'); ('none', 'slight') v. ('some','appreciable','severe+v.severe'); ('none', 'slight','some') v. ('appreciable','severe+v.severe'); ('none','slight','some','appreciable') v. 'severe+v.severe';	10 (with main effects) with one parameter set to zero; 15 (with two 1st order inter.) 2 parameters set to zero.

Table 5.6: Summary of the Statistical Methods used to fit the Regression Models (cont'd)

Model	Method	SAS/STATA Procedure	Health Status score		5-grouped Townsend Disability Scale categories	
			Cut-points	*Number of Parameters fitted*	*Cut-points*	*Number of Parameters fitted*
Constrained Partial Proportional Odds	Weighted least squares	SAS - PROC CATMOD (clogit -response)	('good', 'fair', 'poor') v. 'excellent'; ('fair', 'poor') v. ('excellent', 'good'); 'poor' v. ('excellent', 'good', 'fair').	6 (with main effects); 8 (with interaction).	'none' v.('slight', 'some', 'appreciable', 'severe+v.severe'); ('none', 'slight') v. ('some','appreciable','severe+v.severe'); ('none', 'slight','some') v. ('appreciable','severe+v.severe'); ('none','slight','some','appreciable') v. 'severe+v.severe';	Not fitted
Adjacent Category (constant slopes)	Weighted least squares	SAS - PROC CATMOD (alogit -response)	'good' v. 'excellent'; 'fair' v. 'good'; 'poor' v. 'fair'.	5 (with main effects) 6 (with interaction) (design matrix as for proportional odds)	'none' v. 'slight'; 'slight' v. 'some'; 'some' v. 'appreciable'; 'appreciable' v. 'severe+v.severe'	7 (with main effects); 8 (with interaction) (design matrix as for proportional odds).
Adjacent Category (different slopes)	Weighted least squares	SAS - PROC CATMOD (alogit -response)	'good' v. 'excellent'; 'fair' v. 'good'; 'poor' v. 'fair'.	9 (with main effects) 12 (with interaction) (design matrix as for polytomous model)	'none' v. 'slight'; 'slight' v. 'some'; 'some' v. 'appreciable'; 'appreciable' v. 'severe+v.severe'	16 (with main effects); 20 (with interaction) (design matrix as for polytomous model).
Continuation Ratio (fully constrained)	Weighted least squares	SAS - PROC CATMOD (computed response functions) PROC LOGISTIC (constant odds assumption)	'good' v. 'excellent'; 'fair' v. ('excellent', 'good'); 'poor' v. ('excellent', 'good', 'fair').	5 (with main effects)	'none' v.('slight', 'some', 'appreciable', 'severe+v.severe'); 'slight' v. ('some','appreciable','severe+v.severe'; 'some' v. ('appreciable','severe+v.severe'); ' appreciable' v. 'severe+v.severe'	7 (with main effects); 8 (with interaction)
Continuation Ratio (Binary logistic model)	Maximum Likelihood	SAS - PROC CATMOD (logit -response) PROC LOGISTIC (constant slopes assumption)	(3 separate binary logistic regressions) with multiple testing: 'good' v. 'excellent'; 'fair' v. ('excellent', 'good'); 'poor' v. ('excellent', 'good', 'fair').	Each model: 3 parameters (with main effects) 4 parameters (with interaction)	(4 separate binary logistic regressions): 'none' v.('slight', 'some', 'appreciable', 'severe+v.severe'); 'slight' v. ('some','appreciable','severe+v.severe'; 'some' v. ('appreciable','severe+v.severe'); ' appreciable' v. 'severe+v.severe'	Each model: 4 (with main effects); 5 (with one 1st order inter.).

Table 5.6: Summary of the Statistical Methods used to fit the Regression Models (cont'd)

Model	Method	SAS/STATA Procedure	Health Status score		5-grouped Townsend Disability Scale categories	
			Cut-points	*Number of Parameters fitted*	*Cut-points*	*Number of Parameters fitted*
Continuation Ratio (unconstrained)	Weighted least squares	SAS - PROC CATMOD (computed response functions)	'good' v. 'excellent'; 'fair' v. ('excellent', 'good'); 'poor' v. ('excellent', 'good', 'fair').	9 (with main effects) 12 (with interaction) (design matrix as for polytomous model)	'none' v.('slight', 'some', 'appreciable', 'severe+v.severe'); 'slight' v. ('some','appreciable','severe+v.severe'; 'some' v. ('appreciable','severe+v.severe'); ' appreciable' v. 'severe+v.severe';	16 (with main effects); 20 (with interaction) (design matrix as for polytomous model).
Stereotype Model (with unknown constraints)	Maximum likelihood	SAS - MCLEST (Hendrickx, 2000); STATA - SOREG (Lunt, 2001)	'good' v. 'excellent'; 'fair' v. 'excellent'; 'poor' v. 'excellent'.	7 parameters (with main effects) – with one set to zero 8 parameters (with interaction) – with one set to zero	'none' v. 'severe+v.severe'; 'slight' v. 'severe+v.severe'; 'some' v. 'severe+v.severe'; 'appreciable' v. 'severe+v.severe'	10 (with main effects); 11 (with one 1st order inter.).
Stereotype Model (with known constraints)	Maximum likelihood	SAS - PROC CATMOD (logit -response)	'good' v. 'excellent'; 'fair' v. 'excellent'; 'poor' v. 'excellent'.	5 parameters (with main effects)	'none' v. 'severe+v.severe'; 'slight' v. 'severe+v.severe'; 'some' v. 'severe+v.severe'; 'appreciable' v. 'severe+v.severe'	7 (with main effects); 8 (with one 1st order inter.).

CHAPTER 6 - MODEL CHECKING

6.1 Aims of this Chapter

The objectives of this chapter are to:

(i) assess the overall goodness-of-fit of the model to the data;
(ii) apply residual analysis (if necessary);
(iii) compare different ordinal regression models.

After reviewing the literature, the issues that appear problematic or are not addressed with regards to model checking included:

(a) *the residual analysis of poorly-fitted observations* : This is little explored in the context of ordinal regression models. No statistical methods exist to examine the individual residuals. However, the cell probabilities/frequencies are usually assessed visually (as stated in section 3.6.2).

(b) *The goodness-of-fit of the stereotype model*: The non-linear stereotype model cannot be checked for goodness-of-fit.

(c) *Model assumptions of the adjacent category model*: No method is cited to check the homogeneity of the different slopes adjacent category model.

These issues are addressed in this chapter. The format of this chapter is as follows: in section 6.2 and 6.3 the overall goodness-of-fit and residual analysis (where necessary) of each regression model fitted to the Health Status data and the Townsend Disability Scale data respectively is illustrated. Section 6.4 illustrates how different models (nested and otherwise) are compared. Section 6.5 ends this chapter by highlighting how the above issues have been dealt with.

6.2 Health Status data

6.2.1 Linear Regression Model

(a) Overall goodness-of-fit to the data

The techniques to assess the overall fit of the model to the observed data and assess the individual observations are well established for the linear regression models. Here, the fit of the model, or equivalently, how well the model predicts the dependent variable is examined in two ways:-

1. by considering the proportion of the total sum of squares that can be explained by the regression (R^2). The quantity R is the *multiple correlation coefficient* and R^2 (sums of squares obtained by regression/total sums of squares) is not a measure of the goodness-of-fit of the model, but does give a crude assessment of the overall fit of the model. This statistic ranges from 0 to 1 and generally the larger the statistic the better the model fit.
2. By assessing the Normal plot – whereby the residuals against the standardised normal deviate (normal quantiles) are plotted.

(b) Residual Analysis of individual observations

Due to the limited number of response values and sub-populations, the usual residual plots, namely plots for each covariate and the response against the residuals were not viable. Instead residuals based on the observed and expected response values for the individual observations were obtained. Also, the Normal plot of the residuals was provided and used to isolate outliers and influential observations.

6.2.2 Binary Logistic Regression Models

(a) Overall goodness-of-fit to the data

The individual binary logistic regression models were fitted using the method of maximum likelihood estimation. The overall goodness-of-fit of each of the models to the data was based on the Likelihood Ratio (L.R.) chi-squared statistic given in section 3.30.

(b) Residual Analysis of the individual observations

The residuals based on individual observations for each binary model was assessed using the regression diagnostics listed in section 3.6.2. The following was examined:-

(i) the fit of each observation, by calculating the individual components of the individual components of the Deviance statistic (given in (3.31));

(ii) the measure of influence that was based on the change in deviance goodness-of-fit statistic when an observation was deleted.

6.2.3 Ordinal Regression Models

(a) Overall goodness-of-fit to the data using the Likelihood Ratio Test statistic

The overall goodness-of-fit to the data was assessed using the Likelihood Ratio test statistic (given in section 3.30) for the following models:

- Polytomous Model;
- Different slopes Continuation Ratio Model (using separate binary logistic models);
- Linear Stereotype Model.

The cut-points of the different slopes continuation ratio model were used to fit separate binary logistic regression models. The adjusted Type I error rate (α=0.05/3=0.02) was used when assessing the goodness-of-fit of each binary logistic model.

(b) Overall goodness-of-fit to the data using the Wald goodness-of-fit test statistic

The overall goodness-of-fit to the data was assessed using the Wald goodness-of-fit test statistic (Q_w) for the following models:

- Proportional Odds Model;
- Unconstrained Partial Proportional Odds Models;
- Constrained Partial Proportional Odds Models;

- Adjacent Category Models (different slopes and constant slope);
- Fully constrained and unconstrained Continuation Ratio Model.

The Wald goodness-of-fit test statistic is based on the observed and fitted logits and is given in (A12) in Appendix II. It is distributed as chi-squared for moderately large sample sizes (for instance when $n_{i+} \geq 25$), and its degrees of freedom are equal to the difference between the number of rows of $F(p)$ (or logits) and the number of parameters.

(b) Residual Analysis

The individual residuals were assessed for models where the goodness-of-fit was inadequate. For the Health Status data it was found that models that provided statistically significant goodness-of-fit statistics (indicating lack-of-fit) were:

(i) Proportional Odds;
(ii) Constant Odds Continuation Ratio Model;
(iii) Constant Odds Adjacent Category;
(iv) Linear Stereotype Model.

Models (i), (ii) and (iii) assumed a constant slope over the cut-points and there was evidence of a violation of this model assumption. These models were compared with the different slopes models and there was suggestion that the lack-of-fit was due to the constant slope assumption. As these models could not be used to summarise the results, the need to assess the residuals did not arise.

The assessment of the residuals was only carried out for the stereotype ordinal regression model.

The assessment of the goodness-of-fit of the stereotype model (where the constraints are not known) was not possible. This was because the facility to obtain the predicted cell probabilities is not available in the macros that compute this model (as devised by Hendrickx, (2000)). Also, no facility is available to obtain the predicted logit functions. However, the linear version of the stereotype model (as describe in section 5.2.3.6) provided a means of checking the fit of the model.

The cut-points of the stereotype model could not be used to fit the separate binary logistic regression models for checking poor fit, in the same way as was done for the polytomous model (in Begg and Grey (1984)). This was because the logits of the stereotype model are not the same as those of binary models due to the conditional parameters. In the light of this, the analysis of the residuals was based on the visual assessment of:-

1. residuals from the individual logits (observed and predicted);
2. residuals from the individual cell probabilities (observed and predicted).

The observed and predicted cell probabilities were directly obtained from the *SAS* output and therefore a formulation to compute these was not required.

6.3 Townsend Disability Scale data

6.3.1 Linear Regression Model

The methods used to assess the model checking and residual analysis were the same as those stated for the Health Status data.

6.3.2 Binary Logistic Regression Model

(a) Overall goodness-of-fit to the data

Some of the cell frequencies had sparse data for the Townsend disability score. The overall goodness-of-fit was therefore tested using the Hosmer-Lemeshow statistic.

(b) Residual Analysis of the individual observations

The usual regression diagnostics for the binary logistic regression models (as detailed for the Health Status data) were computed to assess the fit of the individual observations.

6.3.3 Ordinal Regression Models

The overall goodness-of-fit statistics were provided for all models. However, the residual analysis was only carried out if the goodness-of-fit statistic indicated a poor-fit model. Furthermore, for the constant slope models, provided the model assumptions were satisfied, but the goodness-of-fit statistic indicated an inadequate fit, the residuals were assessed.

6.3.3.1 Polytomous Regression Models

(a) The overall goodness-of-fit to the data

The overall fit of the polytomous model was again assessed using the Likelihood Ratio test statistic.

(b) Residual Analysis of the individual observations

The cut-points specified for the polytomous model were made, separating the data into three binary groups. The reason for this is as given by Begg and Gray (1984), who suggest that the fit of the polytomous model can be examined more closely by using regression diagnostics devised for binary logistic regression models (see section 3.6.2). For instance, for an outcome variable with four categories, the fit of the three binary logistics regression models are examined separately and then the results integrated, in a descriptive manner, to make a statement about the fit of the polytomous model.

For our model the following was carried out :-

1. three individual binary logistic regression models were fitted and the parameters from these models compared to those of the polytomous model. As stated in section 3.6.2, Begg and Gray (1984) showed that the estimates of the logistic regression coefficients are usually close to those obtained using the polytomous model and the loss of efficiency is not too great.
2. Each separate binary logistic model was checked for an overall goodness-of-fit using the Hosmer-Lemeshow method (as there was an ordinal covariate – age-group and some of the cell frequencies were ≤ 5). Begg and Gray (1984) did not use multiple testing techniques (i.e. a corrected Type I error rate using

the Bonferroni correction was 0.01 (α=0.05/4)). However, in order to integrate the results of the fit of all three binary logistic regression models to the overall polytomous model, it was considered essential to use the corrected Type I error rate.

3. Where the fit of a binary logistic model was inadequate, the residuals from the individual observations were assessed to explore for outliers or influential observations using the residual diagnostics statistics.

The need to assess the cell probabilities (and logits) of the polytomous model were not considered essential here, as the models based on the binary cut-points provided a reasonable way of assessing the fit and the residuals.

6.3.3.2 Partial Proportional Odds Models

(a) Overall goodness-of-fit to the data

The overall fit of the unconstrained partial proportional odds model to the observed data was assessed using:

1. the Q_w–goodness-of-fit statistic which was based on the logits (observed and expected);
2. the goodness-of-fit statistic as devised by Lipsitz (1996). The need for the use of this statistic arose due to the presence of the low cell frequencies. This statistic is a generalisation of the Hosmer-Lemeshow statistic and is computed when presented with an ordinal outcome (see Appendix III, section 1 for details on computing the statistic).

(b) Residual Analysis of the individual observations

In order to identify the outliers and influential subjects –

1. the individual observed and fitted logit functions were computed and the residuals assessed visually;
2. the individual observed and predicted cell probabilities were computed and the residuals obtained and examined visually.

The expected/predicted cell probabilities ($\hat{\pi}_{ij}$) were not available in *SAS*, so they had to be computed using the logit functions.

Given $f_{i4}(\hat{\pi}) = \ln(\hat{\pi}_{i1} + \hat{\pi}_{i2} + \hat{\pi}_{i3} + \hat{\pi}_{i4} / \hat{\pi}_{i5})$ and $1 = \hat{\pi}_{i1} + \hat{\pi}_{i2} + \hat{\pi}_{i3} + \hat{\pi}_{i4} + \hat{\pi}_{i5}$ it could be derived that $\hat{\pi}_{i5} = \dfrac{1}{e^{f_{i4}(\hat{\pi})} + 1}$. Also given that $f_{i3}(\hat{\pi}) = \ln[(\hat{\pi}_{i1} + \hat{\pi}_{i2} + \hat{\pi}_{i3})/(\hat{\pi}_{i4} + \hat{\pi}_{i5})]$ it can be established that $\hat{\pi}_{i4} = \dfrac{1 - \hat{\pi}_{i5} + \hat{\pi}_{i5} e^{f_{i3}(\hat{\pi})}}{e^{f_{i3}(\hat{\pi})} + 1}$. The $\hat{\pi}_{i1}$ cell probability was easily obtained by re-arranging the equation $f_{i1}(\hat{\pi}) = \ln\left(\dfrac{\hat{\pi}_{i1}}{\hat{\pi}_{i2} + \hat{\pi}_{i3} + \hat{\pi}_{i4} + \hat{\pi}_{i5}}\right)$ to give $\hat{\pi}_{i1} = \dfrac{e^{f_{i1}(\hat{\pi})}}{1 + e^{f_{i1}(\hat{\pi})}}$. From the logit, $f_{i2}(\hat{\pi}) = \ln\left(\dfrac{\hat{\pi}_{i1} + \hat{\pi}_{i2}}{\hat{\pi}_{i3} + \hat{\pi}_{i4} + \hat{\pi}_{i5}}\right)$, the $\hat{\pi}_{i2}$ cell probability can be derived and this took on the form $\hat{\pi}_{i2} = \dfrac{e^{f_{i2}(\hat{\pi})} - \hat{\pi}_{i1} e^{f_{i2}(\hat{\pi})} - \hat{\pi}_{i1}}{1 + e^{f_{i2}(\hat{\pi})}}$, given that $\hat{\pi}_{i1}$ was known. The final cell probability $\hat{\pi}_{i3}$ was obtained when these latter probabilities were known from $1 = \hat{\pi}_{i1} + \hat{\pi}_{i2} + \hat{\pi}_{i3} + \hat{\pi}_{i4} + \hat{\pi}_{i5}$.

6.3.3.3 Constant slope Adjacent Category Models

(a) Overall goodness-of-fit to the data

Although the statistic proposed by Lipsitz (1996) was the correct method to assess the overall goodness-of-fit of the constant slopes adjacent category model, this statistic was difficult to compute. Therefore the overall goodness-of-fit was examined using the Q_w-statistic that was based on the observed and expected logits.

(b) Residual Analysis of the individual observations

The separate binary logit models could not be fitted given the constant slopes adjacent category model. The residuals based on the following were visually assessed:

1. the observed and expected logits;
2. the observed and expected cell frequencies.

The predicted cell probabilities for this model were obtained from the logits functions. Thus let $f_{i1}(\hat{\pi}) = \ln(\hat{\pi}_{i1} / \hat{\pi}_{i2})$, $f_{i2}(\hat{\pi}) = \ln(\hat{\pi}_{i2} / \hat{\pi}_{i3})$, $f_{i3}(\hat{\pi}) = \ln(\hat{\pi}_{i3} / \hat{\pi}_{i4})$ and $f_{i4}(\hat{\pi}) = \ln(\hat{\pi}_{i4} / \hat{\pi}_{i5})$. Then using $1 = \hat{\pi}_{i1} + \hat{\pi}_{i2} + \hat{\pi}_{i3} + \hat{\pi}_{i4} + \hat{\pi}_{i5}$ and the logits were re-arranged in terms of $\hat{\pi}_{i2}$ and it could be derived that

$\hat{\pi}_{i2} = \dfrac{1}{1 + e^{f_{i1}(\hat{\pi})} + e^{-f_{i2}(\hat{\pi})} + e^{-f_{i2}(\hat{\pi})-f_{i3}(\hat{\pi})} + e^{-f_{i2}(\hat{\pi})-f_{i3}(\hat{\pi})-f_{i4}(\hat{\pi})}}$. On obtaining $\hat{\pi}_{i2}$, the remaining probabilities could be derived using $\hat{\pi}_{i1} = \hat{\pi}_{i2} e^{f_{i2}(\hat{\pi})}$, $\hat{\pi}_{i3} = \hat{\pi}_{i2} e^{-f_{i2}(\hat{\pi})}$, $\hat{\pi}_{i4} = \hat{\pi}_{i2} e^{-f_{i2}(\hat{\pi})-f_{i3}(\hat{\pi})}$ and $\hat{\pi}_{i5} = \hat{\pi}_{i2} e^{-f_{i2}(\hat{\pi})-f_{i3}(\hat{\pi})-f_{i4}(\hat{\pi})}$.

6.3.3.4 Different slopes Adjacent Category Model

(a) Overall goodness-of-fit to the data

The overall fit of the different slopes adjacent category model was checked using the Q_w-goodness-of-fit statistic.

(b) Residual Analysis of individual observations

Separate binary logistic regression models were fitted using the cut-points of the separate slopes adjacent category model. Assessing the residuals:

1. the parameters of the separate binary logistic models were compared to those of the different slopes adjacent category model for similarity.
2. Regression diagnostics were used to identify the poorly fitted or influential observations.
3. The results of the binary analyses were integrated to give an overall result of the different slopes adjacent category model.

6.3.3.5 Fully constrained Continuation Ratio Model

(a) Overall goodness-of-fit to the data

The overall goodness-of-fit of this model was assessed in the same was as for the constant slope adjacent category model.

(b) Residual Analysis of the individual observations

The separate binary logit models could not be fitted (as for the constant slopes adjacent category model). The individual observations were assessed using:

1. the observed and expected logits;
2. the observed and expected cell frequencies.

The estimated cell probabilities were obtained from thee predicted response functions. Let $f_{i1}(\hat{\pi}) = \ln(\hat{\pi}_{i1} / \hat{\pi}_{i2} + \hat{\pi}_{i3} + \hat{\pi}_{i4} + \hat{\pi}_{i5})$, $f_{i2}(\hat{\pi}) = \ln(\hat{\pi}_{i2} / \hat{\pi}_{i3} + \hat{\pi}_{i4} + \hat{\pi}_{i5})$, $f_{i3}(\hat{\pi}) = \ln(\hat{\pi}_{i3} / \hat{\pi}_{i4} + \hat{\pi}_{i5})$ and $f_{i4}(\hat{\pi}) = \ln(\hat{\pi}_{i4} / \hat{\pi}_{i5})$. Then using $1 = \hat{\pi}_{i1} + \hat{\pi}_{i2} + \hat{\pi}_{i3} + \hat{\pi}_{i4} + \hat{\pi}_{i5}$ and re-arranging $f_{i1}(\hat{\pi})$ in terms of $\hat{\pi}_{i1}$, it can be derived that

$\hat{\pi}_{i1} = \dfrac{e^{f_{i1}(\hat{\pi})}}{1 + e^{f_{i1}(\hat{\pi})}}$. From $f_{i2}(\hat{\pi}) = In(\hat{\pi}_{i2} /(1 - \hat{\pi}_{i1} - \hat{\pi}_{i2}))$ it can be obtained that

$\hat{\pi}_{i2} = \dfrac{e^{f_{i2}(\hat{\pi})}(1 - \hat{\pi}_{i1})}{1 + e^{f_{i2}(\hat{\pi})}}$. Also the subsequent cell probabilities can be obtained in a similar way

such that $\hat{\pi}_{i3} = \dfrac{e^{f_{i3}(\hat{\pi})}(1 - \hat{\pi}_{i1} - \hat{\pi}_{i2})}{1 + e^{f_{i3}(\hat{\pi})}}$ and $\hat{\pi}_{i4} = \dfrac{e^{f_{i4}(\hat{\pi})}}{1 + e^{f_{i4}(\hat{\pi})}}$. Using these $\hat{\pi}_{i5}$ can be obtained since $\hat{\pi}_{i1} + \hat{\pi}_{i2} + \hat{\pi}_{i3} + \hat{\pi}_{i4} + \hat{\pi}_{i5} = 1$.

6.3.3.6 Different slopes Continuation Ratio Models

(a) Overall goodness-of-fit to the data

Separate binary logistic regression models were fitted using the four cut-points based on the continuation ratio logits and the results integrated to give an overall conclusion of the fit of the different slopes continuation ratio model. The fit of each model was assessed using the Hosmer-Lemeshow test. The adjustment of the Type I error rate (α=0.05/4=0.01) was made for multiple testing.

The fit of unconstrained continuation ratio model was assessed using the Wald residual (Q_w) test statistic.

(b) Residual Analysis of the individual observations

The residual analysis for the different slopes continuation ratio model were examined by fitting separate binary models to the cut-points and then applying the usual regression diagnostics. Thus for both sets of models (i.e. binary logistics and unconstrained models) the same residual analysis was used.

6.3.3.7 Stereotype Model

(a) Overall goodness-of-fit to the data

The overall goodness-of-fit was checked using the Likelihood Ratio test statistic.

(b) Residual Analysis of the individual observations

The residuals for the individual observations were examined visually using:-

3. the observed and fitted logit functions;
4. the individual observed and predicted cell probabilities (as provided by the *SAS* output).

6.4 Comparison of Ordinal Regression Models

Nested Models

The following models were nested:

i. the proportional odds, unconstrained partial proportional odds and the constrained partial proportional odds models;
ii. the fully constrained and different slopes continuation ratio models;
iii. constant slopes and different slopes adjacent category models.

The comparison of the cumulative models in (i) has already been discussed in section 5.2.3.2 (v). Testing the homogeneity in the cut-point specific regression parameters (given a covariate) provided a test of constant slope in the different slopes adjacent category model. This was used to decide whether a constant slopes model was significantly better than a different slopes one. The constant slope assumption for the fully constrained continuation ratio model was checked using the score test (as for the proportional odds model). However, the method used for comparing the adjacent category models could easily be adapted for the continuation ratio models.

Note that although the stereotype model is nested in the polytomous model, given a single covariate, this does not remain the case as covariates increase in the model. The dependency of the parameters in the non-linear stereotype model does not allow this model to be nested in the polytomous given more than one covariate model.

Different Ordinal Regression Models

The Akaike Information Criteria (*AIC*) was used to compare different models from the same data and is defined as $AIC = -2\text{Log}L+2p$, where p is the number of parameters. This criterion adjusted the –2log-likelihood statistics for the number of parameters in the model. The criteria could not be computed for all the models: for the partial proportional odds models, continuation ratio and adjacent category models where the weighted least squares estimation was used, it was not possible to obtain a value for this statistics (as the log-likelihood is not provided). Although the proportional odds model had been fitted using weighted least squares method, it is possible to fit this model in *PROC LOGISTIC* that uses the method of maximum likelihood. Thus, the *AIC* was computed using both dataset for the polytomous, proportional odds and the stereotype models.

6.5 Issues addressed in this Chapter

Issues that are under-researched with regards to the model checking/residual analysis of ordinal regression models have been listed at the beginning of this chapter. Statistically methods used to assess these issues were outlined through the chapter. In particular it should be noted:

(a) the residuals of the logits, in addition to the cell probabilities were used to visually assess the individual observations.

(b) The linear version of the stereotype model was used to examine the goodness-of-fit of the non-linear version. In doing this, the estimates of the ϕ_j parameters are taken as constants and this imposes a limitation in that we are only assessing a single ϕ_j as opposed to a range of them.

(c) The binary analysis for the different slopes continuation ratio model was adjusted for multiple testing to allow an overall conclusion to be drawn for the latter ordinal regression model.

(d) The method for checking the homogeneity in the different slopes adjacent category was similar to that used for checking the proportional odds in the different slopes cumulative logit model.

Table 6.1: Summary of the methods used to assess the overall goodness-of-fit and individual observations of the regression models

Model	Overall Goodness-of-fit		Residual Analysis of Individual Observations	
	Health Status data	***Townsend Disability Scale data***	***Health Status data***	***Townsend Disability Scale data***
Linear Model	a. R^2-test statistic; b. Normal Plots.	a. R^2-test statistic; b. Normal Plots.	-	-
Binary Logistic Regression	Likelihood Ratio χ^2-statistic	Hosmer-Lemeshow test statistic	Regression Residual Diagnostics*	Regression Residual Diagnostics*
Polytomous	Likelihood Ratio χ^2-statistic	Likelihood Ratio χ^2-statistic	-	Separate binary logistic models: a. Check model fit using the Hosmer-Lemeshow test; b. Regression residual Diagnostics*
Proportional odds / Partial Proportional odds	Wald Q_w -test statistic	a. Wald Q_w -test statistic; b. Goodness-of-fit statistic as devised by Lipsitz (1996).	-	Assessing residuals (observed-expected) of: a. Individual logits; b. individual cell probabilities;
Constant slope Adjacent Category	Wald Q_w -test statistic	Wald Q_w -test statistic	-	Assessing residuals (observed-expected) of: a. Individual logits; b. Individual cell probabilities;
Different slopes Adjacent Category	Wald Q_w -test statistic	Wald Q_w -test statistic	-	Separate binary logistic models: a. Check model fit using the Hosmer Lemeshow statistic; b. Regression residual Diagnostics*
Continuation Ratio (separate binary logistic models)	Likelihood Ratio χ^2-statistic	Hosmer-Lemeshow test statistic	-	Separate binary logistic models: a. Check model fit using the Hosmer Lemeshow statistic; b. Regression residual Diagnostics*
Unconstrained Continuation Ratio model	Wald Q_w -test statistic	Wald Q_w -test statistic	-	
Fully constrained Continuation Ratio Model	Wald Q_w -test statistic	Wald Q_w -test statistic	-	Assessing residuals (observed-expected) of: a. Individual logits; b. Individual cell probabilities;
Stereotype	Likelihood Ratio χ^2-statistic	Likelihood Ratio χ^2-statistic	Assessing residuals (observed-expected) of: a. Individual logits; b. individual cell probabilities.	Assessing residuals (observed-expected) of: a. Individual logits; b. individual cell probabilities.

* - where individual observations are examined using the Pearson's chi-squared and deviance statistics (i.e. measures of outliers) and using the change in Pearson's chi-squared statistic (i.e. measure of influence)

CHAPTER 7 - RESULTS AND FINDINGS

7.1 Aims of this Chapter

The primary aim of this chapter is to illustrate the results from the analyses of the two datasets.

The results will be presented in six sections:

Section 7.2 - Statistical modelling;
Section 7.3 - Calculation of the Asymptotic Relative Precision;
Section 7.4 - Interpretation of the results;
Section 7.5 - Comparison of the models;
Section 7.6 – 'Best-fit' models for the two datasets;
Section 7.7 – Summary.

At the end of each section (7.2, 7.3, 7.4 and 7.5) the findings are summarised.

7.2 Statistical modelling

Here the results of the statistical modelling are given. Two different strategies were chosen to include the terms in the models.

(i) *The Health Status data*: Both main effect terms were considered equally important, as one of the objectives was to assess the effect of 'smoke' when fitting/not fitting 'heart attack'. The model assumptions (e.g. the proportional odds, constant slope for the adjacent category/continuation ratio) had to be satisfied by both covariates.

(ii) *The Townsend Disability Scale data*: Attention was given to 'full-time education', as this was the main variable of interest. The model assumptions therefore had to be satisfied by this covariate even though this was not necessarily the case for the adjusting ones (these were 'forced' to satisfy the assumptions).

This section is sub-divided into the following:

- the chosen models;
- goodness-of-fit;
- residual analysis.

7.2.1 Health Status data

There were 12535 subjects who had non-missing health status response and covariate assessments.

7.2.1.1 The chosen models

For all the regression models (with the exception of the binary logistic regressions and the binary logistic regressions based on the continuation ratio cut-points), the main effects models with 'smoke' and 'heart attack' was considered to be the more efficient models, as the interaction terms were statistically non-significant (see Appendix III: section 2). In addition to this, the following points are highlighted:

(a) Binary logistic regression models

For the binary logistic analysis, three models were fitted (as detailed in section 5.2.2). It was found that different models provided a good fit to different cut-points. The main effects model was adequate in terms of fit for the 2nd and 3rd cut-points: ('*fair*', '*poor*') v. ('*excellent*', '*good*') and '*poor*' v. ('*excellent*', '*good*', '*fair*'). However, for the model based on the 1st cut-point (('*good*', '*fair*', '*poor*') versus '*excellent*') the 1st order interaction term had to be fitted (since this term was significant in the model: $\chi_1^2 = 4.72$ with $p = 0.03$) and so the model was saturated. The main effects model for this cut-point was also fitted to compare the results. It was evident from this that, different cut-point specific binary logistic regression models summarised the ordinal data in different ways (with some models having significant interaction terms and others not).

Table 7.1: Assumption of constant slope tested for each ordinal regression model

Model	***Variable***	***Assumption (global/or covariate specific)***	***χ^2- test statistic***	***Degrees of freedom***	***p-value***
Proportional odds		Global assumption	16.21	4	0.003
	smoke		15.51	2	0.0004
	Heart attack		0.58	2	0.74*
Adjacent Category	smoke		11.32	2	0.004
	Heart attack		11.22	2	0.004
Continuation ratio		Global assumption	70.30	4	<0.0001
	smoke		5.98	2	0.05
	Heart attack		64.46	2	<0.0001

* Assumption of constant odds did not hold

(b) Proportional odds model

Table 7.1 illustrates the results from testing the assumption of constant odds. The global score test statistic suggested that the constant slope assumption could not be assumed for both covariates (there are 4 df as an overall test was provided with the two parameters across the (3-1) logits). The proportional odds assumption was tested further using (5.9) and (5.10). It was found that the assumption did not hold for 'smoke' but there were proportional odds for 'heart attack'. As a result the unconstrained and constrained partial proportional odds models were fitted with the assumption of constant odds for 'heart attack' and non-proportionality for 'smoke'. The proportional odds model was not used to summarise the results although the parameters estimates and odds ratios were obtained for comparison with the partial proportional odds models.

(c) Adjacent Category models

From Table 7.1 it is evident that the constant slope assumption did not hold for both covariates. The method of testing the constant slope assumption is given in section 5.2.3.3 (i) (b). This model was not used to summarise the results. However, the parameter estimates and odds ratios were obtained so as that comparison could be made with the different slopes adjacent category model.

(d) Continuation Ratio models

The score test provided evidence of a violation of the parallel slopes assumption over the cut-points for both covariates (see Table 7.1). There was borderline evidence of a constant slope over the cut-points when 'smoke' was fitted on its own and there was evidence that a constant slope could not be assumed for 'heart attack'. The fully constrained model was not considered for summarising the results although the parameters and odds ratios were obtained. Instead attention was focused on the continuation ratio model based the different cut-points (as detailed in section 5.2.3.4 (ii)).

The binary logistic regression analysis based on the continuation ratio cut-points were similar to the above models in (a), in that models based on cut-points 2 and 3, i.e. '*fair*' v. ('*excellent*', '*good*') and '*poor*' v. ('*excellent*', '*good*', '*fair*') had non-significant interaction terms (adjusting for multiple testing, using the corrected Type I error rate of 0.02) and the remaining cut-point specific model was saturated. However, as the results from these three binary logistic models had to be amalgamated and used to provide the overall results for the continuation ratio model, they all had to be consistent in the number of parameters fitted. Therefore these models had to be based either on main effects (with one model providing lack of fit but model parsimony present) or they all had to be saturated models (with two of the models fitted with redundant parameters resulting in no model parsimony but goodness-of-fit achieved). In the end, the decision was to go with the saturated models.

Following these models, the unconstrained continuation ratio model was fitted (using the method described in section 5.2.3.4 (ii) d) and the overall interaction term was not significant (χ_3^2=6.72 with p=0.08). This raised questions regarding the fit of continuation ratio model using the separate binary logistic models. Although in theory the continuation ratio model can be separated into different binary regressions, in practice, there is a chance of spuriously significant effects emerging (despite the application of multiple testing). Also there is the issue of fitting effects into the binary logistic models that do not really need to be fitted for consistency of parameters. This highlights the advantage of fitting the unconstrained continuation ratio model and emphasises the need to fit this model using a simpler way method as was done in this thesis (rather than using the ad hoc method of Cole and Ananth (2001)).

After observing the odds ratios of the unconstrained continuation ratio model, the partially constrained model was considered (details are given in section 5.2.3.4 (iii)).

(e) Stereotype Model

The linear stereotype model was compared to the conditional logistic version. As expected, the –2log-likelihoods for these models were the same confirming that the same models were being fitted (see sections 5.2.3.5 and 5.2.3.6 for methods used to fit the models). The (β_k) regression coefficients were also similar for both models. The parameters $2\beta_1$ and $2\beta_2$ from the linear stereotype model (in Table 7.6) are approximately equal to the parameters β_1 and β_2 from the non-linear stereotype model (using *mclest*). The variation in the estimates and their standard error is as a result of the different methods used (one uses PROC CATMOD the other uses PROC PHREG). This provided one with the evidence that the linear version of the stereotype model was an adequate method for testing the goodness-of-fit of the non-linear version.

The following models were used to summarise the results for the Health Status data:

- Linear Regression Model;
- Binary Logistic Regression Models (with main effects/interaction terms);
- Polytomous Model (with main effects);
- Unconstrained Partial Proportional odds Model (with main effects);
- Constrained Partial Proportional Odds Model (with main effects);
- Different slopes Adjacent Category Model (with main effects);
- Binary Logistic regression models based on the continuation ratio cut-points (with interaction terms);
- Unconstrained/Partially constrained Continuation Ratio Model (main effects);
- Stereotype Model (with main effects).

7.2.1.2 Goodness-of-fit

The goodness-of-fit statistics for the regression models fitted to the data are illustrated in Table 7.2.

(a) Binary Logistic Regression Models

The main effects binary logistic regression model based on the 1st cut-point was the only model that provided an inadequate fit as the interaction term for this model was significant (as specified above and also see Appendix III: section 2). The statistical significance of this latter term were borderline and there is a possibility that this finding may be a spurious, particularly since three separate binary models were fitted to the data.

(b) Linear and Ordinal Regression Models

In addition to the linear regression model, the models that provided inadequate fit to the data were the ones where the cut-point specific regression parameters had certain restrictions imposed on them. The model assumptions for the proportional odds, constant slope adjacent category and constant slope continuation ratio models were violated. It was expected that the goodness-of-fit would be inadequate for these models and this was found to be the case (see Table 7.2). Therefore these latter models were not used to summarise the results and there was no need to assess the goodness-of-fit further.

Note that the main effects unconstrained continuation ratio model provided a good fit, further emphasizing that the use of binary logistic models to summarise the continuation ratio logits should be considered with caution. Also, constraining some parameters to be equal (in the partially constrained continuation ratio model) retained the goodness-of-fit. Since the binary logistic regression models for the continuation ratio logits were saturated, there was no residual variation and the fit was prefect.

Table 7.2: Goodness-of-fit of the main effects regression models fitted using the Health Status data

Model	Goodness-of-fit statistic	Test statistic value	d.f. (no. of logits – no. of parameters)	p-value
Linear	R^{2}†	0.03	-	- *
Binary Logistic (main effects)				
Model for cut-point 1	L.R. χ^2	4.41	1	0.04 *
Model for cut-point 2	L.R. χ^2	0.16	1	0.69
Model for cut-point 3	L.R. χ^2	0.00	1	0.96
Polytomous	L.R. χ^2	6.38	3	0.09
Proportional odds	Wald-Q_w	22.76	7	0.002 *
Unconstrained partial proportional odds	Wald-Q_w	7.14	5	0.20
Constrained partial proportional odds	Wald-Q_w	7.14	6	0.31
Adjacent Category (constant slope)	Wald-Q_w	29.35	7	0.0001 *
Adjacent Category (different slopes)	Wald-Q_w	6.68	3	0.08
Continuation Ratio (fully constrained)	Wald-Q_w	32.42	7	<0.0001*
Continuation Ratio (unconstrained)	Wald-Q_w	6.72	3	0.08
Partially constrained Continuation Ratio	Wald-Q_w	6.95	5	0.22
Linear stereotype model	L.R. χ^2	20.27	7	0.005 *

† - R^2 gives a crude assessment of the fit of the model and is not a goodness-of-fit statistic

*- models indicate lack-of-fit

- For the binary logistic regression models Model1 is based on the cut-point ('*good*', '*fair*', '*poor*') v. '*excellent*'; Model 2 is based on the cut-point ('*fair*', '*poor*') v. ('*excellent*', '*good*'); Model 3 is based on the cut-point '*poor*' v. ('*excellent*', '*good*', '*fair*').
- The binary models based on the different slopes continuation ratio model were saturated.

7.2.1.3 Residual analysis

(a) Linear Regression model

On visual inspection of the residuals, no outliers from the observed and predicted response values were found although the residuals for the '*poor*' category for the non-smokers appeared to be relatively large. The Normal plot - Plot IIIa (in Appendix III) of the residuals failed to identify any outliers/influential observations. This plot is not reliable as the points are in clusters, due to the nature of the data and therefore it is not sensitive enough to highlight the observations that have large residuals and may possibly be outliers.

(b) Ordinal Regression Models

Residual analysis was only computed for the stereotype model.

The residuals of the observed and predicted cell probabilities and response functions for the stereotype model were visually assessed. Generally the residuals from the cell probabilities provided little information regarding outliers or influential observations. Large residuals were apparent for all the logits, for those who had suffered from a heart attack, regardless of whether they smoked or not.

For the binary and stereotype models, there was no one group of subjects that provided poor fit. However, I must emphasise that the inspection of the residuals from the logits for the stereotype model was purely visual.

7.2.2 Townsend Disability Scale data

There were 12434 subjects with complete observations on the Townsend disability score and the three covariates.

7.2.2.1 The chosen models

The forward selection procedure was used to include terms in the model and details of this are given in Appendix III (section 2).

The main effects ('sex', 'age-group' and 'full-time education') model was the best-fit model for the binary analysis (fitted using the method given in the MCR CFAS[1] (1998) study), the constant slope adjacent category model and the constant slope continuation ratio model. For the linear regression and remaining ordinal regression models (with the exception of the unconstrained partial proportional odds model), the main effects model with the interaction of 'age group *x* sex' was the chosen model. The unconstrained partial proportional odds model had the interaction of 'sex *x* full-time education' in addition to the interaction of 'age group *x* sex'. The assumptions for these terms in this model are as stated in 5.3.3.2 (iii).

In addition to the above, the following were noted.

(a) Different slopes ordinal regression models

For the different slopes models, each covariate was fitted separately and its significance assessed in the model. No assumptions were required to be satisfied. For these models, 'age group', 'sex' and 'full-time education' were found to be significant when fitted singularly. As a result initially, the two adjusting covariates were included in the model and then 'full-time education' was added and its effect examined.

(b) Binary Logistic Regression Model

The unadjusted and adjusted effects fitted were corrected for over-dispersion (details of the method used for the model fitting are given in 5.3.2).

(c) Proportional Odds Model

The different slopes cumulative logit model as specified in section 5.3.3.2 (i) was used to determine the proportionality of the covariates. For this model there was evidence of proportional odds for the adjusting covariates when fitted singularly. Allowing for these in the model, 'full-time education' had borderline non-proportional odds (χ_3^2= 8.46 with p =0.04). The worst scenario was taken and it was assumed that the assumption did not hold for the covariate of interest. Therefore the results using the proportional odds model were not summarised (although the parameter estimates and odds ratios were obtained for comparison with the partial proportional odd models).

(d) Constant slope Adjacent Category Model

Of the three covariates fitted singularly, only 'full-time education' demonstrated any evidence of a constant slope. When the adjusting covariates where fitted together (i.e. 'age group' + 'sex'), there was still evidence of heterogeneity in the cut-point parameters (Test of homogeneity: 'sex': χ_3^2= 9.47 with p =0.02 and 'age group': χ_3^2= 32.94 with p <0.001). 'Full-time education' was added into the model, allowing for the adjusting covariates, and the covariate of interest continued to demonstrate homogeneity. Thus, for this model, the adjusting covariates were 'forced' to have a constant slope over the cut-points. The 'full-time education' covariate continued to have homogeneity in the parameter estimates (test for homogeneity for 'full-time education' allowing for constant slopes for the adjusting covariates: χ_3^2= 4.33 with p =0.23). Thus the constant slope adjacent category model was considered, with the assumption of homogeneity in the parameter for all of the three covariates.

(e) Fully constrained Continuation Ratio Model

Each covariate was initially fitted singularly and there was evidence that 'age group' and 'sex' violated the constant slope assumption ('age group': χ_3^2=156.61 with p<0.0001; 'sex': χ_3^2=93.92 with p<0.0001) whilst 'full-time education' was found to satisfy the assumption (χ_3^2=2.88 with p =0.41). The adjusting covariates were constrained to have constant odds and 'full-time education' continued to have homogeneity in its cut-point specific parameters (χ_3^2=3.23 with p=0.38). As a result the fully constrained model was fitted.

In addition to this model the different slopes continuation ratio models were fitted so as to compare results for the different methods.

(f) Different slopes Continuation Ratio Model

For the continuation ratio model where separate binary logistic regression models were fitted (as detailed in section 5.3.3.4 (ii)), multiple testing was used to correct the Type I error rate (corrected α =0.05/4=0.01 (since there were four logistic regressions fitted)). As for the binary logistic regressions, the unconstrained continuation ratio model was fitted with the main effects and the interaction of 'age group *x* sex'.

(g) Stereotype Model

The –2log-likelihood values of the linear version and conditional logistic regression model used to fit the non-linear version were found to be the same, confirming that the same model was being fitted (details of the method are described in section 5.3.3.5 and 5.3.3.6).

The following models were used to summarise the Total Disability Scale data:

- Linear Regression (with interaction 'age group *x* sex');
- Binary logistic (main effects);
- *Different slopes ordinal regression models*
  - Polytomous (with interaction 'age group *x* sex');
  - Adjacent Category (with interaction 'age group *x* sex');
  - Unconstrained Continuation Ratio (with interaction 'age group *x* sex');
  - Binary logistic models based on the continuation ratio (logits with interaction 'age group *x* sex');
  - Linear Stereotype (with interaction 'age group *x* sex');
- *Constant slope ordinal regression models*
  - Constant slope Adjacent Category model (main effects);
  - Fully constrained Continuation Ratio model (main effects);
- *Constant/different slopes ordinal regression models*
  - Unconstrained Partial Proportional Odds (with interactions 'age group *x* sex' and 'sex *x* full-time education').

7.2.2.2 Goodness-of-fit

The goodness-of-fit statistics for all the models suggested lack-of-fit (see Table 7.3).

The following points were noted:

(i) All the regression models provided a poor fit to the data. The reason for the lack of fit was due to the fact that the covariate 'age group' was constrained to take on an ordinal structure (except in the binary logistic model). It was discovered that when identical models were fitted, with 'age group' as a categorical covariate, all the models provided adequate fit to the data, with the exception of the constant slope adjacent category and the fully constrained continuation ratio models. The lack of fit of these latter models was due to the fact that the adjusting covariates had the assumption of constant slopes imposed on them.

(ii) The lack of fit of the constant slope models was due to two factors (a) violation of the constant slope assumption for all covariates and (b) taking 'age group' as ordinal.

(iii) The goodness-of-fit statistic for the unconstrained continuation ratio model should be identical to the sum of the chi-squared values obtained from the individual binary logistic regression models. However, as the goodness-of-fit statistic for the unconstrained continuation ratio model was based on the Wald test statistic it did not provide the summed likelihood ratio chi-squared statistics given by the binary models. The degrees of freedom were the same for both models (as expected). The overall conclusion regarding goodness-of-fit of the different slopes continuation ratio model was that the model did not fit well. The separate binary logistic regression and the unconstrained continuation ratio models provided the evidence for this.

The goodness-of-fit of the unconstrained partial proportional odds was also checked using the Lipsitz's statistic. For this, the disability response categories were given the scores, s_j as 1 –'*none*', 2-'*slight*', 3- '*some*', 4- '*appreciable*' and 5- '*severe + very severe*'. Details of the computation of this statistic are given in the Appendix III section 1. The observed and expected cell probabilities for each subject were used to compute this statistic. There were 12434 subjects each with their own observed and expected probabilities. The expected and observed mean score, i.e. μ_i and $\hat{\mu}_i$ were obtained by multiplying the observed and predicted probabilities by the score (s_j) the subject had obtained for their disability. The predicted mean scores were sorted from the smallest to the largest, and then ten groups (G) were formed: the first nine groups had 1243 subjects and the 10^{th} group had 1247 subjects. Thus a 'group' variable was created with 10 levels; level 1 had the smallest predicted mean scores and the level 10 had the largest predicted mean scores. Then nine dummy or design variables were created using the first group as the referent group. These variables were initially added to the unconstrained partial proportional odds model to form the alternative model (as indicated by model A23 in the Appendix III). The aim was to obtain the parameter estimates $\{\gamma_i; i=1,.. G-1\}$ based on these group variables and then test the hypothesis H_0: $\gamma_1 = ... = \gamma_{G-1}$. The design matrix for this alternative model required fitting 12 covariates (i.e. 9 dummy variables + 3 main effects). Also 51 estimated parameters were required (36 parameters based on the cut-points and the groups and 15 parameters as fitted in the unconstrained partial proportional odds model). This was computationally very intensive, as the design matrix requires sub-populations, formed by nesting the covariates into one another. As a result, the goodness-of-fit statistic was not computed for the unconstrained partial proportional odds, the constant slope adjacent category and the constant slope continuation ratio models. Instead the results from the Wald goodness-of-fit statistic based on the logits were used.

Table 7.3: Goodness-of-fit of the regression models fitted using the Townsend Disability Scale data

Model	**Goodness-of-fit statistic**	**Test statistic value**	**d.f. (no. of logits – no. of parameters)**	**p-value**
Linear	$R^{2\dagger}$	0.22	-	-
Binary Logistic	Hosmer-Lemeshow	22.19	7	0.002
Polytomous	L.R. χ^2	145.19	60	<0.0001
Proportional odds	Wald- Q_w	160.61	72	<0.0001
Unconstrained partial proportional odds	Wald-Q_w	136.73	65	<0.0001
Adjacent Category (constant slope)	Wald-Q_w	219.43	72	<0.0001
Adjacent Category (different slopes)	Wald-Q_w	146.52	60	<0.0001
Fully constrained continuation ratio	Wald-Q_w	451.52	72	<0.0001
Binary Logistic regression models based on continuation ratios	L.R. χ^2			
Model for cut-pt. 1		62.85	15	<0.0001
Model for cut-pt. 2		51.80	15	<0.0001
Model for cut-pt. 3		59.65	15	<0.0001
Model for cut-pt. 4		18.81	15	0.22
Continuation Ratio (unconstrained)	Wald-Q_w	184.19	60	<0.0001
Linear stereotype model	L.R. χ^2	193.31	72	<0.0001

† - R^2 gives a crude assessment of the fit of the model and is not a goodness-of-fit statistic

For the binary logistic regression models Model1 is based on the cut-point '*none*' v. '*slight*', '*some*', '*appreciable*', '*severe + very severe*'); Model 2 is based on the cut-point ('*none*', '*slight*') v. ('*some*', '*appreciable*', '*severe + very severe*'); Model 3 is based on cut-point ('*none*', '*slight*', '*some*') v. ('*appreciable*', '*severe + very severe*'); Model 4 is based on the cut-point ('*none*', '*slight*', '*some*', '*appreciable*') v. '*severe + very severe*'.

7.2.2.3 Residual analysis

(a) Linear Regression Model

The residuals from the observed and fitted values illustrated no extreme outliers, although some large residual values were apparent (e.g. males in the '*severe + very severe*' category given age group under 70 years with level of education 13 or more years and females in the '*severe + very severe*' category given age group under 70 years with 13 or more years full-time education). These are not shown up in the Normal plot (Plot IIIb (in Appendix III)) and again there was indication that the Normal plots were not sensitive enough to highlight the outliers/influential observations.

(b) Binary Logistic Regression model

The regression diagnostics (as given in section 6.3.2) showed that the fit of the binary logistic regression model became less accurate as the scale index increased (i.e. there were large residuals towards the latter end of the scale then towards the 'no disability' end). This implied that the lack of fit of this model was largely due to the subjects who had a greater degree of disability compared to those with little or no disability.

(c) Different slopes Ordinal Regression Models

The polytomous model, different slopes adjacent category model and the unconstrained continuation ratio model were fitted using separate binary logistic regressions for the purpose of residual analysis (see section 6.3.3 for details).

i. Polytomous Model: Table IIIa in Appendix III illustrates the parameter estimates from the fit of the binary analysis based on the cut-points of the polytomous model. These parameter estimates were based on the maximum likelihood method. There were some similarities (approx. to 2 d.p.) in the parameter estimates of 'full-time education' in Table IIIa (Appendix III) and those of the polytomous model in Table 7.15. The Hosmer-Lemeshow statistics for the binary models confirmed that the polytomous model provided a poor-fit. The results from the binary analysis were related back to the polytomous model. The diagnostic statistics showed the fit was worse for the '*none*' v. '*severe + very severe*' cut-point and it improved as the logits increased. The residual deviance was relatively large for the observations where the cell frequencies were less than or equal to 5. For the first three logits, although no one group of subjects had lack of fit, there was indication that males and females who were < 70 years of age regardless of their years of education fitted the model inadequately.

ii. Different slopes adjacent category model: Table IIIb in Appendix III provides the estimates of the binary logistic regression models that were fitted using the adjacent category logits. These parameter estimates, which were based on maximum likelihood estimation, were noticeably different to those provided by the different slopes adjacent category model (estimated using weighted least squares (see Table 7.15)). The difference in the estimates could be explained by the fact that different estimation methods were used. The residual analysis using the regression diagnostics was carried out. The Hosmer-Lemeshow test statistics indicated an adequate fit for all binary cut-point specific models with the exception of the one based on the cut-point '*some*' v. '*severe* + *very severe*'. The results from the diagnostic analysis were integrated to given overall results of the residual analysis for the different slopes adjacent category model. The deviance residual chi-square statistic values for some of the residuals from this model were large, particularly for a number of observations where the cell frequency was five or less.

iii. Unconstrained continuation ratio model: There is indication from the binary analyses that all the models, with the exception of the one based on the final cut-point, provided an inadequate fit (see Table 7.3). The lack of fit for the model based on cut-point 1 is largely contributed by females, aged 85 or more years in the '*none*' disability group for both 'full-time education' groups. For the model based on cut-point 2, this response category is dropped and the fit remains poor. Again the lack of fit is largely contributed by the same sub-population in the '*slight*' disability grouping. The '*slight*' disability group is dropped from the model based on cut-point 3 and although the influence of the latter group of subjects is reduced there is still enough influence in the data to provide a lack of fit for the binary model. Thus, the overall results from the residual analysis of the unconstrained model indicated that females aged 85 or over years were an influential group leading to an inadequate fitting model.

The residual analysis of the stereotype model was not computed using binary logistic regressions. Instead, the logit functions and cell probabilities were assessed visually. The logits functions (predicted and observed) with their residuals illustrated a potential outlier: females aged <70 years with 13 or more years full-time education. The individual probabilities were also examined and there was no indication of any influential cell frequencies or outliers. However, for females aged <70 years with 13 or more years education the cell frequency of the referent category '*excellent*' was ≤ 5.

(d) Constant slope Ordinal Regression Models

The residuals of the proportional odds model were not provided, as these models were not used to summarise the results.

The residuals from the cell frequencies and logits for the constant slope adjacent category model and the fully constrained continuation ratio model were assessed visually (see section 6.3.3 for details).

i. Constant slope adjacent category model: The logit (predicted and observed) residuals were computed for the individual sub-populations. There was a possible outlier – females aged <70 years with 13 or more years full-time education in logit 4: ('*appreciable*' v. '*severe + very severe*'). The predicted cell probabilities were derived using the predicted logit values and there was no systematic pattern in the residuals and there were no apparent outliers.

ii. Fully constrained continuation ratio model: The residuals from the logits functions (predicted and observed) illustrated some noticeably large values. In particular, females aged <70 years with 13 or more years full-time education were identified as one of the observations with large residual values.

(e) Constant/Different slopes Ordinal Regression Model

For the unconstrained partial proportional odds model, the residuals from the cell probabilities were not very informative. The expected and observed logit functions values were examined visually and there was a potential outlier, namely females with 13 or more years full-time education in the <70 years age group occurring in the ('*none*', '*slight*', '*some*', '*appreciable*') v. '*severe+ very severe*' disability group (i.e. logit 4). The cell count for the '*severe+ very severe*' category for this group of subjects was less than or equal to 5. There were a few relatively large residual values for the remaining row population logit functions; for instance the log odds for (i) females with less than 13 years full-time education and aged $\geq$ 85 years when comparing '*none*' v. ('*slight*', '*some*', '*appreciable*', '*severe+ very severe*') disability (i.e. logit 1) and (ii) males with $\geq$ 13 years full-time education and aged < 70 years when comparing ('*none*', '*slight*', '*some*') v. ('*appreciable*', '*severe+ very severe*') disability (i.e. logit 3).

7.2.3 Findings of the statistical modelling

From the above results the following are highlighted.

Linear models

The assumptions of the linear regression models cannot be checked (as mentioned in section 5.2.1) and the residual analysis is also difficult to carry out due to the nature of ordinal data.

In section 5.2.1, I showed that the assumptions of the linear regression models could not be checked, due to the limited number of response categories and covariate sub-populations. Here it was discovered that the residual plots that are often used when data have been analysed using linear regression models were not appropriate for ordinal outcomes. The residual analysis based on the Normal plots was not sensitive enough to detect large residuals or outliers due to the reduced number of y observations (i.e. the ordinal categories) and sub-populations.

Binary Logistic Regression models

There is evidence to suggest that the binary logistic regression analysis is not a very reliable method for fitting ordinal data compared to ordinal regression models. (i) Different types of effects may be required for the cut-point specific binary models. There is a possibility of spuriously significant effects emerging due to multiple number of models fitted to the cut-points; (ii) extra parameters, which are redundant, may be required to be fitted in order to obtain consistency (for all the binary models), so as that an overall conclusion can be drawn and related to the ordinal scale. However, this does not result in efficiency or model parsimony.

- For the binary logistic regression models and the binary logistic regression models based on the continuation ratio cut-points using the Health Status data, there is a chance that the interaction terms may be spuriously significant. All the regression models (with the exception of the binary analyses) have significant main effects (using the Health Status dataset). The data in Table 4.1 suggest that the proportion of smokers and non-smokers are quite different within the two 'heart attack' categories for '*excellent*' health status. There are more smokers than non-smokers in the heart attack group and there are fewer smokers than non-smokers in the no heart attack group. This would suggest that any log odds based on this response category is likely to demonstrate an interaction effect. For the remaining response categories, however, there is consistency in the number of smokers/ non-smokers within the heart attack groups indicating similarity in effect across the smokers/non-smokers groups for the two 'heart attack' categories.

- Also, it was expected that the binary logistic analysis based on the continuation ratio cut-points and the unconstrained continuation ratio model would give the same number and type of effects. However, this was not the case for the Health Status data. The interaction term was found to be significant, when fitting the separate binary model based on cut-point 1. The overall interaction of 'heart attack' and 'smoke' was non-significant for the unconstrained continuation ratio model. This would imply that although a cut-point specific interaction term is significant when fitted in a single binary model, its effect is 'diluted' and diminishes when allowing for the entire response categories in an ordinal regression model (where other interaction effects are present). This then results in an overall non-significant interaction term.

- The binary analysis for the different slopes continuation ratio model (given the Health Status data) had to allow for the extra parameters for consistency over all the cut-points and therefore this model required fitting effects that did not necessarily need to be fitted. The similarity that should occur in both types of the different slopes models was illustrated by the Townsend Disability Scale data where the number and type of parameters for the unconstrained continuation ratio model and the binary logistic regression models were identical.

<u>*Ordinal Regression Models*</u>

Ordinal regression models are more efficient in fitting ordinal data compared to the linear and binary models.

- The unconstrained continuation ratio model was much more efficient with regards to the parameters fitted compared to the binary logistic regression models (as indicated by the models fitted using the Health Status data). The problems of fitting extra parameters or multiple testing do not occur for ordinal regression models as they occur for binary models.

However, constant slope ordinal regression models may be of little use as the model assumptions are too stringent.

- It was found from the goodness-of-fit statistics (using the Health Status data) that sacrificing model parsimony and fitting a large number of parameters (i.e. different slopes models) provided good fit. In contrast, fitting constant slope models where the model was parsimonious provided a poor fit model and the chances of the model assumptions being satisfied were highly unlikely. This suggested that the constant slope models (i.e. proportional odds, constant slope continuation ratio, constant slope adjacent category) and the linear stereotype model were of little use due to the stringent assumptions. Furthermore, there was evidence from the datasets, that if the constant slope assumption fails for one model, it may fail for all.

- Provided the covariate of interest satisfies the model assumption, then the other covariates can be 'forced' to satisfy the model assumptions (as illustrated using the Townsend Disability Scale data). If, however, the covariate of interest does not satisfy homogeneity in the cut-point specific parameters, then it is of little value in knowing how the adjusting covariates are behaving. In these circumstances the global χ^2-score test of homogeneity is of little use.

The partially constrained models (partial proportional odds and the partially constrained continuation ratio models) are more efficient than the different slopes and the constant slope ordinal regression models.

- The partially constrained continuation ratio model allowed the possibility of constraining some parameters to be equal (increasing model parsimony) without sacrificing the goodness-of-fit (for the Health Status data). Similarly, the unconstrained partial proportional odds model was a good-fit model and had reduced number of parameters compared to the different slopes cumulative model.

Evidence suggested that the Lipsitz's (1996) procedure for testing the goodness-of-fit of an ordinal regression model given an ordinal covariate was of little value.

- The methods based on the maximum likelihood (Likelihood Ratio test) and weighted least squares (Wald goodness-of-fit statistic) which use the logit functions are ample for testing the goodness-of-fit of the ordinal regression models. The Lipsitz's (1996) procedure was found to be too computationally intensive.

If at all possible, the residual analysis should be carried out using separate binary analyses for the ordinal regression models (as was done by Begg and Gray, (1984)). Otherwise, the residual analysis has to be based on the visual assessment of cell probabilities/logit functions and this does not provide one with strong statistical evidence to identify outliers/influential observations.

- The residual analysis was carried out using residual diagnostics, where ordinal models could be fitted using binary analysis and this appears to be a satisfactory way of analysing the residuals of the observations. However, for some models one cannot do this (such as the stereotype model for the Health Status data and the constant slope adjacent category, fully constrained continuation ratio models using the Townsend Disability Scale data). Here the assessment of the residuals is based on the logit/cell probabilities and this is visual and not supported by any statistics. Therefore one has to be cautious as regards the conclusions drawn from this analysis.

- The residuals from the cell probabilities appeared to be less sensitive in picking out the influential observations and there is evidence to suggest that the residuals from the logits need to be used in conjunction with those of the cell probabilities.

There is some evidence to suggest that if an outlier is very influential then it will affect the fit of all the ordinal regression models.

- There were no common outliers for the models fitted using the Health Status data, whereas for the Townsend Disability Scale data, there was a common outlier for almost all the regression models fitted (with the exception of the binary logistic regression model and the unconstrained continuation ratio model): females aged <70 years with 13 or more years education. This indicates that one may not necessarily be able to identify a common influential/outlier for all the models, unless it is very influential.

Given the data are skewed, such that there is sparse information in one of the extreme categories of the scale, then these data are problematic if (i) the cells where there are sparse data have been used as independent data points, as opposed to being cumulated with other *cells, when computing the logit functions, and (ii) when forming the logit functions, if there is a marked difference in the marginal probabilities of the denominator and in the numerator of the logit function.*

- There was only one subject who was a female, with 13 or more years education, aged 70 years with '*severe + very severe*' disability (see Table 4.2). Lack-of-fit resulted when this category was used singularly in the formation of the logits. No problems in terms of fitting were encountered when this category was amalgamated with others (e.g when forming cumulative probabilities) and logits were formed. From the residual analysis (based on diagnostic statistics) of the polytomous and the unconstrained continuation ratio models, all the logits provided poor fit, with the exception of the final one (where '*severe + very severe*' was used an a single category in the formation of the log odds). For the latter logit, the marginal probabilities (used in the numerator and denominator) of the logits were relatively similar. Also from the diagnostic statistics provided using the different slopes adjacent category model, the relative difference for the marginal probabilities used to form logit 3 was bigger than that of any of the log odds, leading to a poorly predicted log odds. For the models where the residual analysis was based on the visual assessment of the logits, (i.e. the partial proportional odds models and the stereotype model), similar results were found although this could not be based on any statistical evidence.

Stereotype Ordinal Regression Model

The stereotype model was initially devised to maintain the goodness-of-fit of the polytomous model, and at the same time allow for model parsimony. There is little evidence that this model has done this based on the data presented in this thesis.

- The results of the Health Status data indicate that the goodness-of-fit is more adequate when ordering constraints are not imposed (polytomous model) compared to when they are (stereotype model). This would suggest that use of the stereotype model defeats the aim of achieving model parsimony without sacrificing the goodness-of-fit.

It would appear that the linear version of the stereotype model can be adequately used to check the residuals of the non-linear.

- The linear form of the stereotype model can be used to check for goodness-of-fit of the non-linear satisfactorily, since the –2log-likelihoods and the parameter estimates (see Table 7.6) for both models were identical.

Based on the findings, I believe that the use of the stereotype model is likely to be limited in the application of quality of life research.

- The indistinguishability aspect of the stereotype model is important as it leads to model simplification. However, this involves comparing the –2-log-likelihoods of the models where indistinguishability was imposed and where it was not. The macros devised by Lunt (2001) allow these stereotype models to be computed, but his macros are not compatible with the bootstrapping techniques of *Stata* whereby the models could be compared and indistinguishability tested. This would suggest further work was needed in this area.

- The ordering parameters, ϕ_j are estimated parameters from the model and therefore one should compute the 95% confidence intervals for these. This was done for both datasets. The confidence intervals computed using the Health Status data illustrated no overlapping, implying that the ϕ_js were indeed ordered and therefore the *y*-response was ordinal with respect to the covariates. For the Townsend Disability Scale data, the 95% confidence intervals of these parameters overlapped. Also, there was an indication that the point estimates of ϕ_j were not monotonic confirming an absence of ordinality in the *y*-response categories (with respect to the covariates). Thus the 95% confidence intervals of the ordering parameters provide additional information about the precision of the point estimates.

- Various authors have highlighted the use of the 2-dimensional stereotype model. However, in the context of the quality of life data presented here, its use was limited (given both types of response variables). Furthermore, even if its application proved to be relevant, the question regarding whether the *y*-response was ordered with respect to the covariates would still hold.

- Hendrickx's (2000) macros and Lunt's (2001) macros cope well with several covariates and interaction terms; one does not encounter the same problems as presented by the other ordinal regression models, where as covariates increase, the fit of the model becomes increasing complex (in particular, fitting the design matrix).

However, bootstrapping was very computationally intensive and became even more so, when interaction terms were fitted.

7.3 Calculation of the Asymptotic Relative Precision (ARP)

7.3.1 Health Status data

The Asymptotic Relative Precision (ARP) could not be computed for the partial proportional odds model, as the unadjusted model fit was not possible. Also, the ARP could not be computed for the linear stereotype model as the ordering constraint parameters differed for the unadjusted and adjusted 'smoke' effect, and therefore the two parameter estimates were not comparable. Although the model assumptions did not hold for the constant slope models (using the Health Status data), the ARPs were still computed for these, for comparison with the ARPs of the different slopes models.

The ARP for each model is displayed in Table 7.4. These statistics were derived from the standard errors of the unadjusted and adjusted 'smoke' parameter estimates (as given Tables 7.5 and 7.6). Note that the adjusted smoke effect was based on the saturated binary logistic regression models for the continuation ratio model. These are displayed in Table 7.11.

For the binary models based on the 1st cut-point, the ARP is much smaller when the interaction term is fitted compared to the main effects model. This suggests that the standard error of 'smoke' is inflated as more covariates are added into the model resulting in a smaller ARP. This is also the case for the continuation ratio model, where different binary logistic models were used.

For the different slopes ordinal regression models, namely the polytomous, adjacent category and unconstrained continuation ratio models the ARPs are very close to 1 for all the cut-points, implying that the variability for the 'smoke' covariate is little affected when 'heart attack' is added into the model. For the constant slope models, namely the proportional odds and the adjacent category, again the ARPs are close to 1. There is indication here, that the precision of an adjusted covariate is affected by the number of covariates added into the model and not by the type of parameters fitted (whether constant or different slopes).

7.3.2 Findings of the Asymptotic Relative Precision

The ordinal regression models are as efficient as the binary logistic regression model, in terms of the ARP. Also the loss of precision in the adjusting covariates is not affected by model assumptions (whether constant or different slopes), but rather by the number of covariates present in the model.

Table 7.4: Asymptotic Relative Precisions (ARPs) for the regression models using the Health Status data

Model	**ARP for model based on cut-pt. 1***	**ARP for model based on cut-pt. 2***	**ARP for model based on cut-pt. 3***
Linear		Not applicable	
Binary Logistic (main effects)	0.9932	0.9757	0.9911
Binary Logistic (saturated)	0.2361		
Polytomous	1.000	0.9829	0.9843
Proportional Odds		0.9907	
Unconstrained partial proportional odds		-	
Constrained partial proportional odds		-	
Adjacent Category (constant slope)		0.9716	
Adjacent Category (different slopes)	1	0.9857	0.9959
Fully constrained continuation ratio model		0.9785	
Continuation Ratio (based on binary logistic models)	0.2277	0.4075	0.6292
Continuation Ratio (unconstrained)	1	0.9776	0.9911
Stereotype		-	

*** Refer to Table 5.6 for the cut-points fitted for each model**

- The ARP for the partial proportional odds models for the single 'smoke' covariate could not be computed as the number of parameters required to fit the model exceeded the number of logits.
- The ARP is not applicable to the stereotype model (linear or non-linear).

Table 7.5: Unadjusted parameter estimates (and standard errors) for the binary and ordinal regression models for the 'smoke' covariate using the Health Status data

	Cut-point 1*		Cut-point 2*		Cut-point 3*	
	Estimate	***Standard Error***	***Estimate***	***Standard Error***	***Estimate***	***Standard Error***
Model						
Binary Logistic	0.0618	0.0294	0.0606	0.0242	0.2260	0.0445
Polytomous	0.0421	0.0313	0.0497	0.0346	0.2616	0.0505
Proportional odds	0.0737	0.0213	0.0737	0.0213	0.0737	0.0213
Unconstrained Partial Proportional Odds	-	-	-	-	-	-
Constrained Partial Proportional Odds	-	-	-	-	-	-
Adjacent Category (constant slope)	0.0567	0.0138	0.0567	0.0138	0.0567	0.0138
Adjacent Category (different slopes)	0.0421	0.0313	0.00758	0.0277	0.212	0.0483
Fully constrained continuation Ratio	0.0623	0.0183	0.0623	0.0183	0.0623	0.0183
Continuation Ratio (based on binary logistic regression)	0.0421	0.0313	0.0196	0.0263	0.2260	0.0445
Continuation Ratio (unconstrained)	0.0421	0.0313	0.0196	0.0263	0.2260	0.0445
Stereotype Model (non-linear)	-	-	-	-	-	-

*** Refer to Table 5.6 for the cut-points fitted for each model**

- The 'smoke' covariate could not be fitted for the partial proportional odds models and therefore no estimates have been provided.
- Also no estimate for the 'smoke' covariate is provided for the stereotype model (linear or non-linear) due to the conditioning of the ordering constraint parameters

Table 7.6: Adjusted parameter estimates (and standard errors) for the binary logistic and ordinal regression models of the 'smoke' and 'heart attack' covariates using the Health Status data

	All cut-points		Cut-point 1*						Cut-point 2*				Cut-point 3*			
	Smoke		Smoke		Heart Attack		Smoke *x* Heart Attack		Smoke		Heart Attack		Smoke		Heart Attack	
Model	***Estimate***	***s.e.***	***Estimate***	***s.e***	***Estimate***	***s.e.***	***Estimate***	***s.e.***	***Estimate***	***s.e.***	***Estimate***	***s.e.***	***Estimate***	***s.e.***	***Estimate***	***s.e.***
Binary Logistic-																
Model 1 (saturated)	-	-	-0.054	0.0605	0.4451	0.0605	-0.1313	0.0605	0.0621	0.0245	0.5174	0.0298	0.2281	0.0447	0.4832	0.048
Model 1 (main effects)	-	-	0.0623	0.0295	0.5230	0.0512	-	-	0.0621	0.0245	0.5174	0.0298	0.2281	0.0447	0.4832	0.048
Polytomous	-	-	0.0419	0.0313	0.3098	0.0542	-	-	0.0524	0.0349	0.7197	0.0541	0.2656	0.0509	0.8955	0.066
Different slopes Cumulative	-	-	0.0611	0.0295	0.5104	0.0512	-	-	0.0652	0.0244	0.5187	0.0298	0.2296	0.0447	0.4828	0.048
Proportional odds			0.0771	0.0214	0.5111	0.0277	-	-	0.0771	0.0214	0.5111	0.0277	0.0771	0.0214	0.5111	0.027
Unconstrained Partial Proportional Odds	0.0609	0.0295	0	-	0.5115	0.0277	-	-	0.0041	0.0314	0.5115	0.0277	0.1691	0.0503	0.5115	0.027
Constrained Partial Proportional Odds	0.0608	0.0210	0.00423	0.001	0.5115	0.0277	-	-	0.00423	0.001	0.5115	0.0277	0.0042	0.001	0.5115	0.0277
Adjacent Category (constant slopes)	-	-	0.0574	0.0140	0.3284	0.0185	-	-	0.0574	0.0140	0.3284	0.0185	0.0574	0.0140	0.3284	0.0185
Adjacent Category (different slopes)	-	-	0.0413	0.0313	0.3013	0.0544	-	-	0.00878	0.0279	0.4090	0.0334	0.2129	0.0484	0.1754	0.0514
Fully constrained Continuation Ratio model			0.0626	0.0185	0.4480	0.0238			0.0626	0.0185	0.4480	0.0238	0.0626	0.0185	0.4480	0.0238
Continuation Ratio (unconstrained)	-	-	0.0414	0.0313	0.3008	0.0545	-	-	0.0215	0.0266	0.4839	0.0317	0.2281	0.0447	0.4833	0.0484
Continuation Ratio (Partially constrained)	-	-	0.0298	0.0203	0.3013	0.0545	-	-	0.0298	0.0203	0.4837	0.0265	0.2281	0.0447	0.4837	0.0265
Stereotype model (non-linear) using mclest	-	-	0.2761	0.0766	1.8239	0.1000	-	-	0.2761	0.0766	1.8239	0.1000	0.2761	0.0766	1.8239	0.1000
Stereotype model (linear)	-	-	0.1275	0.0353	0.9459	0.0514	-	-	0.1275	0.0353	0.9459	0.0514	0.1275	0.0353	0.9459	0.0514
Stereotype model (non-linear) using the bootstrapping	-	-	0.2703	0.0805	1.8468	0.1346	-	-	0.2703	0.0805	1.8468	0.1346	0.2703	0.0805	1.8468	0.1346

*** Refer to Table 5.6 for the cut-points fitted for each model**

- The parameter estimated for the unconstrained partial proportional odds model are (β_1) as given under the 'all cut-points' column, and (γ_j) as given under the cut-point specific columns.
- The ϕ_j parameters for the stereotype model (non-linear –estimated using *mclest* (Henrickx (2000)) were ϕ_2=0.3370; ϕ_3=0.7766; ϕ_4=1.000.
- The ϕ_j parameters for the stereotype model (non-linear –estimated using the bootstrap technique) were ϕ_2=0.3407; ϕ_3=0.777; ϕ_4=1.000.
- The parameter estimates for the continuation ratio model based on the separate binary logistic regressions are given in Table 7.11.

7.4 Interpretation of the results

7.4.1 Health Status data

7.4.1.1 Linear Regression Model

The parameter estimates for the main effects linear regression model are given in Table 7.7. From above, it is clear that the linear regression model is inadequate for summarising ordinal data. Generally, it is not good statistical practice to interpret parameters where the models are found to be inadequate. However, in this thesis this was done to purely illustrate the differences and drawbacks of the linear models in relation to the other models. The β parameter estimate for the adjusted 'smoke' covariate indicates that a decrease of 0.07 in the health status score is evident for every unit increase in 'smoke'. Also, the β parameter estimates for the 'heart attack' covariate indicates that for every unit increase, there is a decrease of 0.44 in the health status score. Due to the ordinal nature of the quality of life scale and discrete covariates presented, these results are a simplification of the truth. The general indication is that moving from the 'smokers' to the 'non-smokers', health status gets better (i.e. a decrease in the numerical score of the scale indicates an increase in the health status category). Also moving from those who have suffered from a heart attack to those subjects who have not, there is evidence again that health status gets better. Furthermore, the rate at which health status gets better is higher for those who have suffered/not suffered a heart attack then those who smoke/not smoke.

Table 7.7: Parameter estimates (and standard errors) for the linear regression model using the Health Status data

Variable	Parameter Estimate	Standard error	t-test value	p-value
Smoke	-0.0726	0.0182	-3.99	<0.0001
Heart attack	-0.4400	0.0234	-18.77	<0.0001

The parameter estimates from the main effects model were used to compute the expected mean health status scores and these are given in Table 7.8. The latter is a single summary measure over all the response categories and appears to provide the same conclusion for all sub-populations when relating the results back to the ordinal response categories. For instance, for respondents who have suffered from a heart attack and smoke, their expected Health Status score was 1.62 (i.e. somewhere between health status categories '*good*' and '*fair*'). Similarly, for those who have suffered from a heart attack and are non-smokers their expected Health Status score was 1.56 (i.e. again somewhere between health status categories '*good*' and '*fair*'). The same conclusion can be drawn from the remaining two sub-populations, where the mean response falls between categories '*good*' and '*fair*'. The conclusion then is that all subjects have mean health status response of somewhere between '*good*' and '*fair*'.

Table 7.8: The observed and expected adjusted means for each sub-population using the Health Status data

Heart attack	**Smoke**	**Observed Mean**	**Adjusted Expected Mean**	**N**
	Yes	1.62	1.63	243
Yes	*No*	1.56	1.56	1045
	Yes	1.19	1.19	2119
No	*No*	1.12	1.12	9128

7.4.1.2 Binary Logistic Regression Models

The adjusted maximum likelihood parameter estimates (and their standard errors) of the separate binary logistic regression models are given in Table 7.6.

The odds ratios and their 95% confidence intervals are given in Tables 7.9 and 7.10.

Although the main effect model (based on the 1^{st} cut-point) provided an inadequate fit, this model was used to interpret the results. The reason for this was the results were much simpler in interpretation and also, in terms of the parameters fitted, were consistent with the other binary logistic models.

Main effects: The odds ratios based on all three main effect binary logistic models are quite similar for the 'heart attack' covariate. Generally there is indication that the adjusted odds of having a worse form of health status is approximately three times that of a better form of health status for those who had suffered from a heart attack (compare to those who have not). The adjusted odds of ('*good*', '*fair*', '*poor*') health as opposed to '*excellent*' health or ('*fair*', '*poor*') as opposed to ('*good*', '*excellent*') health are almost identical for both those who smoke compared to those who do not. The adjusted odds of '*poor*' as opposed to ('*fair*', '*good*', '*excellent*') is approximately 1.6 times for those who smoke compared to the non-smokers.

Interaction term: For the interaction model based on the 1st cut-point, there is indication that among the non-smokers, the adjusted odds of ('*good*', '*fair*', '*poor*') health was approximately three time that of '*excellent*' health for those who had suffered from a heart attack compared to those who had not. However, for the smokers the odds were much less (approximately twice) for those who had had a heart attack compared to those who had not.

Also, among those who had not had a heart attack, the odds of ('*good*', '*fair*', '*poor*') health were approximately identical to that of '*excellent*' health for those who were smokers compared to the non-smokers. Among those who had suffered from a heart attack, the odds were 0.7 for the smokers compared to the non-smokers.

This latter finding is somewhat unexpected, as there is indication those who suffered from a heart attack and smoke are more likely to have improved health (i.e. fall in the '*excellent*' category) compared to those who suffered from a heart attack and are non-smokers. There may be some underlying factors (e.g. history of smoking or other factors such as level of exercise etc.) that may contribute to the explanation of this finding, but this is beyond the scope of this thesis. For our purpose, it is important to note the finding and assess it relevance in the context of the modelling techniques.

7.4.1.3 Polytomous Model

The adjusted maximum likelihood parameter estimates (and their standard errors) for the main effects polytomous model are illustrated in Table 7.6.

The odds ratios and their 95% confidence intervals are given in Tables 7.9 and 7.10.

For both covariates, the odds ratios were found to increase monotonically over the cut-points.

Table 7.9: Adjusted odds ratios and their 95% confidence intervals for the comparison of 'smokers v. non-smokers' using the Health Status data

		Cut-point 1*		Cut-point 2*		Cut-point 3*	
Model		***Odds Ratio***	***95% C.I. of Odds Ratio***	***Odds Ratio***	***95% C.I. of Odds Ratio***	***Odds Ratio***	***95% C.I. of Odds Ratio***
Binary Logistic- MODEL 1 (interaction)	Heart attack	0.69	(0.44, 1.09)				
	No heart attack	1.17	(1.03, 1.32)				
Binary logistic (main effects)		1.13	(1.00, 1.25)	1.13	(1.03, 1.25)	1.58	(1.32, 1.88)
Polytomous		1.09	(0.96, 1.28)	1.11	(0.97, 1.27)	1.70	(1.39, 2.08)
Proportional odds		1.17	(1.01, 1.29)	1.17	(1.01, 1.29)	1.17	(1.01, 1.29)
Different slopes Cumulative		1.13	(1.01, 1.29)	1.14	(1.04, 1.25)	1.58	(1.33, 1.89)
Unconstrained Partial Proportional Odds		1.13	(1.01, 1.27)	1.14	(1.03, 1.25)	1.58	(1.33, 1.89)
Constrained Partial Proportional Odds		1.13	(1.04, 1.23)	1.14	(1.05, 1.24)	1.58	(1.33,1.88)
Constant slope Adjacent category		1.12	(1.06, 1.85)	1.12	(1.06, 1.85)	1.12	(1.06, 1.85)
Adjacent category (different slopes)		1.09	(0.96, 1.23)	1.02	(0.91, 1.14)	1.53	(1.27, 1.85)
Fully constrained Continuation Ratio		1.13	(1.05, 1.22)	1.13	(1.05, 1.22)	1.13	(1.05, 1.22)
Continuation Ratio (based on the binary logistic regression models)	Heart attack	0.57	(0.34, 0.94)	1.08	(0.80, 1.47)	1.56	(1.05, 2.31)
	No heart attack	1.13	(1.00, 1.28)	1.04	(0.93, 1.16)	1.58	(1.30, 1.92)
Continuation Ratio (unconstrained)		1.09	(0.96, 1.23)	1.04	(0.94, 1.16)	1.58	(1.32, 1.88)
Continuation (partially constrained)		1.06	(0.98, 1.15)	1.06	(0.98, 1.15)	1.57	(1.32, 1.88)
Stereotype Model (non-linear)		1.10	(1.04, 1.16)	1.23	(1.10, 1.38)	1.30	(1.12, 1.53)

- **Refer to Table 5.6 for the cut-points fitted for each model**

Table 7.10: Adjusted odds ratios and their 95% confidence intervals for the comparison of 'heart attack v. no heart attack' using the Health Status data

		Cut-point 1*		Cut-point 2*		Cut-point 3*	
Model		***Odds Ratio***	***95% C.I. of Odds Ratio***	***Odds Ratio***	***95% C.I. of Odds Ratio***	***Odds Ratio***	***95% C.I. of Odds Ratio***
Binary Logistic- MODEL 1 (interaction)	Smokers	1.87	(1.24, 2.84)	-	-	-	-
	Non-smokers	3.17	(2.52, 3.99)	-	-	-	-
Binary logistic (main effects)		2.85	(2.33, 3.48)	2.81	(2.50, 3.16)	2.63	(2.17, 3.18)
Polytomous		1.86	(1.50, 2.30)	4.21	(3.41, 5.22)	6.00	(1.61, 7.79)
Proportional odds		2.78	(2.49, 3.10)	2.78	(2.49, 3.10)	2.78	(2.49, 3.10)
Different slopes cumulative		2.78	(2.27, 3.39)	2.82	(2.51, 3.17)	2.63	(2.17, 3.17)
Unconstrained Partial Proportional Odds		2.78	(2.50, 3.10)	2.78	(2.50, 3.10)	2.78	(2.50, 3.10)
Constrained Partial Proportional Odds		2.78	(2.50, 3.10)	2.78	(2.50, 3.10)	2.78	(2.50, 3.10)
Constant slope Adjacent category		1.93	(1.79, 2.07)	1.93	(1.79, 2.07)	1.93	(1.79, 2.07)
Adjacent Category (different slopes)		1.83	(1.48, 2.26)	2.27	(1.99, 2.58)	1.42	(1.16, 1.74)
Fully constrained Continuation ratio model		2.45	(2.23, 2.70)	2.45	(2.23, 2.70)	2.45	(2.23, 2.70)
Continuation Ratio (based on the binary logistic regression models)	Smokers	1.08	(0.68, 1.70)	2.73	(2.04, 3.65)	2.60	(1.78, 3.81)
	Non-smokers	2.62	(1.66, 2.70)	2.61	(2.28, 3.00)	2.63	(2.12, 3.28)
Continuation Ratio (Unconstrained)		1.82	(1.47, 2.25)	2.63	(2.32, 2.98)	2.63	(2.17, 3.18)
Continuation Ratio (partially constrained)		1.82	(1.48, 2.26)	2.63	(2.37, 2.92)	2.63	(2.37, 2.92)
Stereotype Model (non-linear)		1.88	(1.55, 2.28)	4.19	(3.32, 5.29)	6.34	(4.87, 8.25)

*** Refer to Table 5.6 for the cut-points fitted for each model**

(a) Smokers versus non-smokers

For the smokers (versus the non-smokers), the adjusted odds of '*good*' or '*fair*' health were identical to '*excellent*' health. However, for this group of subjects the adjusted odds of '*poor*' health was almost twice that of '*excellent*' health.

(b) Heart attack versus no heart attack

The adjusted odds of '*good*' health were approximately twice that of '*excellent*' health for those who suffered from a heart attack (compared to those who had not). The adjusted odds of '*fair*' health were four times that of '*excellent*' health for those who had suffered from a heart attack (compared to those who had not), and the adjusted odds of '*poor*' health were six times that of '*excellent*' health for the latter group of subjects.

The implication here is that there if proportional odds holds (as for the 'heart attack' covariate), then the polytomous model will provide odds ratios that increase monotonically (as for the smoke covariate). However, if the proportional odds assumption is found not to hold, then although the polytomous model provides odds ratios that increase monotonically, they do not vary substantially over the cut-points.

The fact that the odds ratios increase monotonically for the polytomous model, given that the proportional odds assumption holds has been proved in Appendix IV.

7.4.1.4 Cumulative Logit Models

(i) Different slopes Cumulative Model

The adjusted weighted least square parameter estimates for the main effects different slopes cumulative model are illustrated in Table 7.6.

The odds ratios and their 95% confidence intervals are given in Tables 7.9 and 7.10.

The adjusted odds ratios for those who were smokers compared to the non-smokers clearly differ between cut-point 1 (('*good*', '*fair*', '*poor*') versus '*excellent*') and cut-point 2 (('*fair*', '*poor*') as opposed to ('*excellent*', '*good*')) compared to cut-point 3 '*poor*' versus ('*excellent*', '*good*', '*fair*'). This would suggest non-proportionality. Also there is not much difference in the cut-point specific odds ratios of 'heart attack', possibly indicating proportional odds.

(ii) Unconstrained Partial Proportional Odds Model

The weighted least square adjusted parameter estimates for the unconstrained partial proportional odds model for both covariates are illustrated in Table 7.6.

The odds ratios and their 95% confidence intervals are given in Tables 7.9 and 7.10.

(a) Smokers versus non-smokers

The odds of (‘*good*’, ’*fair*’, ‘*poor*’) health are almost identical to ‘*excellent*’ health for smokers compared to non-smokers. The odds of (’*fair*’, ‘*poor*’) health are again very similar to (‘*excellent*’, ‘*good*’) health when comparing the smokers with the non-smokers. The odds of ‘*poor*’ health are 1.6 times that of (‘*excellent*’, ‘*good*’, ‘*fair*’) health for the smokers compared to the non-smokers.

(b) Heart attack versus no heart attack

For those who have had a heart attack compared to those who have not, the odds of having a worse form of health are approximately three times that of a better form of health.

The odds ratios for ‘smoke’ based on the first two cut-points are very similar and the difference in the cut-point specific odds ratios emerges at the final cut-point (see Table 7.9). It is at this cut-point that the proportional odds assumption has been violated.

(iii) Constrained Partial Proportional Odds Model

The adjusted weighted least square parameter estimates for the constrained partial proportional odds model are illustrated in Table 7.6.

As the constraints for this model were based on the unconstrained partial proportional odds model, the odds ratios were identical to those computed for this latter model (see Table 7.9 and 7.10).

7.4.1.5 Different slopes Adjacent Category Model

The weighted least parameter estimates with their standard errors for the main parameters are as given in Tables 7.5.These parameter estimates were used to obtain the odds ratios and their 95% confidence intervals and the latter are displayed in Tables 7.9 and 7.10.

(a) Smokers versus non-smokers

The adjusted odds of '*good*' health as opposed to '*excellent*' health or '*fair*' health as opposed to '*good*' health were almost identical for the smokers versus the non-smokers. However, the adjusted odds of '*poor*' health are approximately 1.5 times that of '*fair*' health for the smokers (versus the non-smokers).

(a) Heart attack versus no heart attack

The adjusted odds of '*good*' health as opposed to '*excellent*' health or '*fair*' health as opposed to '*good*' health were almost twice for those who had suffered from a heart attack compared to those who had not. The odds of '*poor*' health is approximately 1.4 times that of '*fair*' health for those who had suffered from a heart attack compared to those who had not.

The odds ratios for both covariates fluctuated in different ways over the cut-points, implying that the behaviour of these covariates with respect to the categories was very different.

7.4.1.6 Continuation Ratio Model

(i) Continuation Ratio Model based on the binary logistic regression models

The maximum likelihood parameter estimates and their standard errors for the saturated binary logistic regression models based on the continuation ratio logits are presented in Table 7.11.

Table 7.11: Parameter estimates and their standard errors for the individual binary logistic regression models based on the Continuation Ratio model using the Health Status data

Variable	**Estimate and s.e.**	**'good' v. 'excellent'**	**'fair' v. ('good', 'excellent')**	**'poor' v.('fair', 'good' v. 'excellent')**
Smoke	Parameter estimate	-0.1072	0.0298	0.2261
	s.e.	0.0656	0.0412	0.0561
Heart attack	Parameter estimate	0.2056	0.4908	0.4816
	s.e.	0.0656	0.0412	0.0561
Heart attack.smoke	Parameter estimate	-0.1691	0.0109	-0.0033
	s.e.	0.0656	0.0412	0.0561

For each cut-point specific model, the presence of the interaction term signifies that an odds ratio exists when comparing smokers with non-smokers for each level of 'heart attack' and when comparing heart attack versus no heart attack within each level of the 'smoke' covariate.

The odds ratios and their 95% confidence intervals are detailed in Table 7.9 and 7.10.

(a) Smokers versus non-smokers

Within the no heart attack group: The odds ratios were not markedly different over the cut-points, when comparing the smoker and non-smokers groups.

Within the heart attack group: The difference emerges in the heart attack group. The odds of '*good*' health were 0.5 times that of '*excellent*' health for the smokers compared to the non-smokers. The scenario here is similar to that of the binary logistic regression models reported in section 7.4.1.2. The unexpected finding, that those who had suffered from a heart attack and smoke are less likely to have a '*good*' health status as opposed to '*excellent*', compared to those who have suffered from a heart attack and are non-smokers, may be explained using similar arguments specified for the binary logistic regression models.

(b) Heart attack versus no heart attack

Within the non-smokers group: the odds of a given health status was approximately three times that of a 'better' state of health (as represented by grouped categories) for those who have had a heart attack as opposed to not having had one.

Within the smokers group: the odds of having '*fair*' health was approximately three times that of ('*good*', '*excellent*') health and the odds of '*poor*' health also three times that of ('*fair*', '*good*', '*excellent*') health for those who had had a heart attack (compared to those who had not). The difference emerges when comparing the odds of '*good*' versus '*excellent*' health for the two heart attack groups. There is indication that within the smokers group, the odds of '*good*' health status is equally likely as '*excellent*' health status for those who have had a heart attack (compared to those who have not).

(ii) Unconstrained Continuation Ratio Model

The weighted least squares parameter estimates for the unconstrained continuation ratio model are detailed in Table 7.6.

The adjusted odds ratios and their 95% confidence intervals are given in Tables 7.9 and 7.10.

(a) Smokers versus non-smokers

The adjusted odds of '*good*' health are almost identical to '*excellent*' health for smokers (compared to non-smokers). Likewise, the adjusted odds of '*fair*' health are almost identical to ('*excellent*', '*good*') health for the same group of people. The adjusted odds of '*poor*' health are approximately 1.5 times that of ('*excellent*', '*good*', '*fair*') health status for the smokers (compared to the non-smokers). So there is little discrimination when comparing '*excellent*', '*good*' and '*fair*' health status categories for the smokers (compared to the non-smokers). The difference merges when comparing the '*poor*' health status with these categories.

(b) Heart attack versus no heart attack

Identical odds ratios for the 'heart attack' covariate were provided for cut-points 2 and 3. The odds of '*good*' health were approximately twice that of '*excellent*' health for those who suffered from a heart attack (as opposed to suffering from a heart attack).

(iii) Partially constrained Continuation Ratio Model

The parameter of the covariates from the unconstrained continuation ratio model, indicate that some may be constrained to be the same. The test of homogeneity revealed that for the 'smoke' covariate, the regression coefficients for cut-points 1 and 2 could be considered as being similar (test of homogeneity in the parameters: χ_1^2=0.23 with p =0.63). There was evidence that the regression coefficients for cut-points 2 and 3 of 'heart attack' could be assumed to be homogenous (test of homogeneity in the parameter estimates: χ_1^2=6.27 with p =0.01). The partially constrained continuation ratio model was then fitted.

(a) Smokers versus non-smokers

This partially constrained continuation ratio model fitted the data well (see Table 7.2). Also it provided parameter estimates (see Table 7.6) and odds ratios (and 95% confidence intervals) that were the same for the first two cut-points for the smokers against the non-smokers (see Table 7.9). Thus the odds of '*good*' health were approximately identical to '*excellent*' health for the smokers (as opposed to the non-smokers). Also the odds of '*fair*' health were approximately the same as ('*excellent*', '*good*') health for the smokers. The odds of '*poor*' health were just over 1.5 times that of ('*excellent*', '*good*', '*fair*') health for the smokers compared to the non-smokers.

(b) Heart attack versus no heart attack

The odds ratios were interpreted in a similar way to the unconstrained continuation ratio model.

7.4.1.7 Stereotype Model

Mclest provided the parameter estimates (note that the macros do not provide the standard error of the ϕ_j j=1, 2, 3) as displayed in Table 7.6. The valid parameter estimates were obtained using bootstrapping. These are also displayed in Table 7.6 (note that there are no standard errors for ϕ_1 and ϕ_4 as there are constrained to 0 and 1 respectively). The valid ϕ_j (j= 2, 3, 4) parameters were ordered, indicating that one can assume there is an ordinal relationship of the response categories with respect to the covariates. The effect of the log odds for any two levels of the outcome is proportional (as signified by the β_k parameters). The difference between the ϕ_j (j=1, 2, 3) parameters provides an indication of how the log odds are affected by the covariates. Thus the difference between any two adjacent ϕ_j parameters was approximately the same (see Table 7.6) indicating the log odds between two outcomes were influenced equally by the covariates.

The valid estimate of the log odds ratios as provided by the bootstrapping technique are displayed in Table 7.9 and 7.10. The interpretation of the odds ratios (and their 95% confidence intervals) was similar to that of the polytomous model.

7.4.2 Townsend Disability Scale data

7.4.2.1 Linear Regression Model

(a) Interaction term

The significant interaction of 'age group *x* sex' suggested that the average predicted effect of the Townsend disability score was not consistent for males within each age group and for females within each age group. Since, results are focused only on the 'full-time education' covariate, this interaction does not contribute to the interpretation of the results.

(b) Interpreting the parameter estimates

As for the Health Status data, the linear model fitted here was not adequate and therefore the interpretation of the parameters was illustrated purely to further emphasis the drawbacks of this model. The adjusted regression coefficient for 'full-time education' from the linear regression model was –0.29 (with s.e. = 0.04). This indicates that there is a decrease in the Townsend disability score of approximately 0.3 for unit increase in 'full-time education'. A significant difference in the two full-time education groups was noted with regards to the Townsend Disability categories (t-test statistic = -8.11 on 1 df.; $p < 0.0001$). In other words as one moves from the < 13 years to ≥ 13 years full-time education categories, the Townsend disability score is found to fall (i.e. disability gets better).

Table 7.12 illustrates the expected mean Townsend disability score for each sub-population. For all age groups within the male/female categories, the adjusted mean Townsend disability score is higher for those with less than 13 years education compared to those with 13 or more years full-time education. However, in terms of the ordinal categories, the mean score falls in the same category for both those in the < 13 and 13 or more years full-time education indicating no change from one category to another (since there is only a change of approximately 0.3 between the mean score in both groups of full-time education).

Table 7.12: The expected and observed Townsend disability score means from the adjusted linear regression analysis

Sex	**F-T education**	**Age-group**	**Observed Mean**	**Expected Mean**
Male	*<13*	*<70*	1.65	1.54
Male	*<13*	*70-74*	1.79	1.91
Male	*<13*	*75-79*	2.19	2.29
Male	*<13*	*80-84*	2.61	2.66
Male	*<13*	*≥85*	3.42	3.03
Male	*≥13*	*<70*	1.55	1.25
Male	*≥13*	*70-74*	1.60	1.62
Male	*≥13*	*75-79*	1.94	1.99
Male	*≥13*	*80-84*	2.32	2.37
Male	*≥13*	*≥85*	2.67	2.74
Female	*<13*	*<70*	1.98	1.82
Female	*<13*	*70-74*	2.21	2.30
Female	*<13*	*75-79*	2.64	2.79
Female	*<13*	*80-84*	3.25	3.27
Female	*<13*	*≥85*	3.95	3.75
Female	*≥13*	*<70*	1.59	1.53
Female	*≥13*	*70-74*	1.96	2.01
Female	*≥13*	*75-79*	2.30	2.49
Female	*≥13*	*80-84*	2.92	2.97
Female	*≥13*	*≥85*	3.62	3.46

7.4.2.2 Binary Logistic Regression Model

The regression parameters for the main effects binary logistic regression (similar to the one fitted in the MRC CFAS paper (MRC CFAS[1], 1998)) were obtained and the parameter of interest was that of 'full-time education'. The adjusted parameter estimate is as illustrated in Table 7.13 and was similar to that provided in the MRC CFAS[1] study (1998) publication where nine adjusting covariates were used.

The interpretation of this parameter was however unclear in the publication and therefore an attempt was made to clarify this.

Table 7.13. Parameter estimates for the binary logistic regression model using the Townsend disability score for full-time education (unadjusted and adjusted effect)

Full-time education	***Unadjusted Effect***		***Adjusted for sex and age***	
	Estimate	***s.e***	***Estimate***	***s.e.***
(four categories as in the (MRC CFAS[1], 1998))				
f-t educ. (<9 v.9 years)	0.5541	0.0370	0.3023	0.0361
f-t educ. (10-12 v. 9 years)	-0.1627	0.0281	-0.0785	0.0266
f-t educ. (>=13 v. 9 years)	-0.4064	0.0412	-0.3369	0.0386
(two categories – as in the binary logistic model fitted in this thesis)				
f-t educ (<13 v. >=13 years)	0.2126	0.0271	0.2083	0.0252

Each subject had a disability score, and this score was calculated as a proportion of the total score. One is effectively looking at the chance of a point (where a subject could have up to 18 points) being allocated when one computes the proportion. This was further clarified by using the *SAS* output as produced in the 'Response Profile Table'. From this output, the table illustrated the 'Binary outcome event total frequency =45640' and 'Binary outcome non-event total frequency=178172'. Using the dataset, 4479 subjects scored '0' for their Townsend disability score, 1489 subjects scored '1', 1243 subjects scored 2 and so on. In the analysis, the number based on 'event' was taken as (0x4479)+(1x1489)+(2x1243)... (18x133)=45640. This number reflects the total number of points given out for all subjects, where no points given indicate no disability and 18 points indicate severe disability. The 'non-event' count was obtained by totalling the number of scores that could be possibly allocated (18x12434) minus the number that were allocated (45640) given a total of 178172. Thus the logit was based on 'event/no event', where 'event' was the total number of points given out for all subjects and 'no event' was the total number of points not allocated. The log odds were based on the probability of being allocated a single point (1/18) as opposed to not being allocated a point.

In the context of the scale, an allocation of a point implies disability. However, the disability scale is a severity scale, where '0' implies no disability and '18' implies very severe disability. The concept of the *severity* of the scale is lost when analysing the data using the method in the publication.

Bearing in mind that the data are fitted poorly by the model, particularly towards the end of the scale, the odds of a point allocated towards the disability score (implying that disability gets worse) was 0.7 times for those with 13 or more years full-time education compared to those with less than 13 years full-time education (see Table 7.14).

Table 7.14: The odds ratios and 95% confidence intervals for full-time education using the binary logistic regression model for the Townsend Disability Scale data

Full-time education	**Unadjusted Effect**		**adjusted for age group and sex**	
	Odds ratios	***95% C.I. of odds ratio***	***Odds ratios***	***95% C.I of odds ratio***
(four categories as in the (MRC CFAS[1], 1998))				
<9	1.71	(1.56, 1.89)	1.21	(1.10, 1.33)
9	1		1	
10-12	0.84	(0.78, 0.90)	0.83	(0.77, 0.88)
>=13	0.66	(0.59, 0.73)	0.64	(0.58, 0.71)
(two categories – as in the binary logistic model fitted in this thesis)				
<13 years	1		1	
>= 13 years	0.65	(0.59, 0.73)	0.66	(0.60, 0.73)

The results as they stand in Table 7.14 are not comparable with those provided by the ordinal regression models. The results given by the ordinal regression models assess the odds of having some form of less disability compared to a more severe form for those with less than 13 years full-time compared to 13 or more years full-time education. The reason for summarising the results in this way was that it was important to keep the referent category as '*severe + very severe*'. This would then allow one to assess the models in the presence of sparse data, as these data were most predominant in this latter response category.

The same odds ratios (and 95% confidence intervals) as provided in Table 7.14 are obtained when one computes the odds of no allocation of a disability score (as opposed to one allocated) for those with < 13 years full-time education compared to 13 or more years education. The interpretation of the results is such that we are now assessing the odds of the disability staying as it is, as opposed to it getting worse for the two groups of 'full-time education'. This would allow the comparison of the binary and ordinal regression analysis.

7.4.2.3 Polytomous Regression Model

(a) Interaction term

The interaction term 'age group *x* sex' suggests that for each cut-point, the effect of the response was different for every level of age group for the male and female categories. As 'full-time education' was of interest here, the comparison of < 13 and ≥ 13 years was not affected by this interaction term. (Note: that an interaction effect at a particular cut-point may not be significant but the overall interaction term may be significant due to other highly significant effects).

(b) Interpreting the parameters

The maximum likelihood parameter estimates and standard errors for 'full-time education' for the polytomous model with the first order interaction term 'age group *x* sex' are given in Table 7.15.

The adjusted odds ratios and their 95% confidence intervals for 'full-time education' are illustrated in Table 7.16. The adjusted odds of '*no*' disability was 0.4 times that of '*severe+ very severe*' disability for those with < 13 years compared to those with ≥ 13 years full-time education. The odds was found to increase over the cut-points, with the odds of '*appreciable*' disability being very close to '*severe + very severe*' disability for those with less compared to more full-time education.

7.4.2.4 Unconstrained Partial Proportional odds Model

(a) Interaction terms

The interaction term of 'age group *x* sex' was automatically assumed to satisfy the proportional odds assumption, as the main effect covariates were assumed to have proportional odds. For a given cut-point the relationship of 'sex' and the log odds changed across the levels of 'age group'. Furthermore, for any sex/age-group combination, the log odds were constant over all the cut-points.

The interaction term 'sex *x* full-time education' did not satisfy the assumption of proportional odds. The log odds of < 13 years full-time education compared to 13 or more years full-time education were different for males and females for any one cut-point. For any sex/full-time education combination the log odds were allowed to vary over the cut-points.

(b) Interpreting the parameter estimates

The parameter estimates and their standard errors are shown in Table 7.15.

The odds ratios and their 95% confidence intervals are illustrated in Table 7.16.

Within the male sex group: The adjusted odds ratios were found to decrease over the cut-points. The odds of '*none*' disability was 0.8 times that of ('*slight*', '*some*', '*appreciable*', '*severe+ very severe*') disability for those with less education compared to more education. This implied that there was not much difference in the two groups of education when comparing no disability with the other grouped disabilities. As the cut-points increase and the categories are amalgamated, the numerator of the log odds ratios is represented by disability groupings that are 'less severe' and the denominator is represented by disability groupings that are comparatively 'more severe'. The difference in terms of effect is large for these groupings for the odds ratios based on the cut-point ('*none*', '*slight*', '*some*', '*appreciable*') versus '*severe+ very severe*' then for the odds ratio based on the cut-point '*none*' versus ('*slight*', '*some*', '*appreciable*', '*severe+ very severe*'). As a result, the odds of having some form of '*less severe*' disability as opposed to '*severe + very severe*' disability is much more in favour of those with more education than those with <13 years full-time education.

Within the female sex group: The adjusted odd ratios were found to drop over the cut-points and then increased for the final cut-point. Thus, the odds of '*none*' disability was 0.7 times that of ('*slight*', '*some*', '*appreciable*', '*severe+ very severe*') disability for females with < 13 years education compared to those with ≥ 13 years full-time education. The odds of ('*none*', '*slight*') compared to ('*some*', '*appreciable*', '*severe + very severe*') were fractionally smaller (0.6) for those with less education compared to those with more education. The odds ratio continued to decrease for the next cut-point and then marginally increased for the final cut-point.

Females, compared to males, were marginally less likely to have '*none*' disability compared to ('*slight*', '*some*', '*appreciable*', '*severe + very severe*') disability and ('*none*', '*slight*') disability compared to ('*some*', '*appreciable*', '*severe+ very severe*') disability. Females, compared to males, were marginally more likely to have ('*none*', '*slight*', '*some*') than ('*appreciable*', '*severe* + *very severe*') disability and ('*none*', '*slight*', '*some*', '*appreciable*') than '*severe* + *very severe*' disability.

7.4.2.5 Adjacent Category Models

(i) Constant slope Adjacent Category Model

(a) Interaction term

The interaction of 'age group *x* sex' for the constant slope adjacent category model implied that the effect of the response given males and females was different within each age group category, but was the same over all the cut-points.

(b) Interpretation of the parameters

The weighted least square parameter estimates for the constant slope adjacent category model are given in Table 7.15.

The adjusted odds of falling in a given disability category as opposed to falling into a disability category that is adjacent to it and more severe, was 0.8 for those with less than 13 years compared to 13 or more years full-time education (see Table 7.16).

(ii) Different slopes Adjacent Category Model

(a) Interaction term

The interpretation of the interaction term for this model was exactly the same as that for the polytomous model (except that different type of logits were used for the two models).

(b) Interpreting the parameter estimates

The parameter estimates obtained from the weighted least squares analysis are displayed in Table 7.15.

The adjusted odds ratios for 'full-time education' are displayed in Table 7.16. In brief, the adjusted odds ratios are found to fluctuate over the cut-points.

The adjusted odds of '*none*' disability compared to '*slight*' disability were 0.9 and were very similar to the adjusted odds of '*slight*' disability compared to '*some*' disability for those with less than 13 years full-time education compared to 13 or more years. This implied that there was not much difference when comparing the 'better' disability categories for the two 'full-time education' groups. The adjusted odds of '*some*' disability compared to '*appreciable*' fell to 0.7 and then increased for the final cut-point '*appreciable*' compared to '*severe+ very severe*' to 0.8 for those with less education compared to more education.

7.4.2.6 Continuation Ratio Models

(i) Fully constrained Continuation Ratio Model

The parameter estimates are given in Table 7.15. The constant odds ratio (as provided in Table 7.16) illustrates that subjects with a given form of disability are 0.7 times likely to have a form worse than it, if they have less than 13 years education compared to if they have 13 or more years education.

(ii) Model based on the separate binary logistic regression models

(a) Interaction term

The presence of the interaction of 'age group *x* sex' for a given binary logistic regression model indicates that the effect of the response within each sex group varied over the levels of age. This interaction has no effect on the 'full-time education' covariate.

(b) Interpretation of the parameters

The parameter estimates and their standard errors were computed from the weighted least squares analysis and are given in Table 7.15.

The adjusted odds ratios and their 95% confidence intervals were obtained and are listed in Table 7.16. The adjusted odds ratios were found to decrease over the cut-points and the odds ratio based on the final cut-points was relatively larger. The adjusted odds of '*none*' disability as opposed to ('*slight*', '*some*', '*appreciable*', '*severe+ very severe*') were very similar to the odds of '*slight*' disability as opposed to ('*some*', '*appreciable*', '*severe+ very severe*'). This implied there was little difference in the 'less severe' forms of disability when compared to the 'more severe' forms for the two groups of education. Also, the odds of a 'less severe' form of disability, albeit small, were in favour of those with more education. The odds of '*some*' disability are just over half those of ('*appreciable*', '*severe+ very severe*') disability for those with less education (compared to those with more education). This would imply that '*some*' disability category is a good discriminator between the 'less severe' and 'more severe' forms of disability compared to the '*none*' and '*slight*' categories. The odds of '*appreciable*' disability are similar to that of '*severe+ very severe*' for the two groups of education.

(iii) Unconstrained Continuation Ratio Model

(a) Interaction term

The interpretation of the interaction of 'age group *x* sex' was similar to that of the polytomous model.

(b) The interpretation of the parameter estimates

The parameter estimates obtained using weighted least squares are detailed in Table 7.15. Note that the estimates differ slightly to those from the separate binary logistic regression models in (i) above. This is because the estimates from the latter model were obtained using maximum likelihood estimation.

The odds ratios and their 95% confidence intervals were identical to those given by the separate binary logistic regression models (displayed in Table 7.16).

7.4.2.7 Stereotype Model

(a) Interaction term

The interpretation of the interaction term is similar to that presented for the polytomous model.

(b) Interpretation of the parameters

The standard error of the parameter estimates obtained when fitting the conditional logits model for the non-linear stereotype model were conditional and therefore could not be used for reporting or inference purposes. Bootstrapping was carried out on a 100 samples (with replacement) to obtain the corrected standard errors of the log odds ratios. The results from the bootstrapping analysis are presented in Table 7.15. Note that the standard errors of the conditional model are dependent and therefore under-estimated. The log odds $\phi_j\beta_k$ were obtained and subsequently the odds ratios and their 95% confidence intervals were derived (see Table 7.16). The odds of '*none*' disability was approximately 0.5 times that of '*severe+ very severe*' disability for those with less than 13 years compared to 13 or more years full-time education. Although the ϕ_j constraints do not increase monotonically over the cut-points, the odds ratios are found to increase in this fashion and the odds of less disability remain in favour of those with more than 13 years full-time education.

Table 7.15: Parameter estimates and their standard errors for each ordinal regression model given the Townsend Disability Scale data

		Cut-point 1*		Cut-point 2*		Cut-point 3*		Cut-point 4*	
Model	**f-t education parameters**	***Parameter***	***s.e.***	***Parameter***	***s.e.***	***Parameter***	***s.e.***	***Parameter***	***s.e.***
Polytomous	β_{j3}	-0.4548	0.0658	-0.3751	0.0673	-0.3098	0.0674	-0.1026	0.0794
Proportional Odds	β_3	-0.2094	0.0287	-0.2094	0.0287	-0.2094	0.0287	-0.2094	0.0287
Unconstrained Partial Proportional odds	Main effect β_3	-0.3442	0.6310	-0.3442	0.6310	-0.3442	0.6310	-0.3442	0.6310
	γ_{j3}	0	-	0.0078	0.0485	0.1279	0.0611	0.1679	0.0655
	Interaction effect (with sex) β_6	-0.0960	0.0414	-0.0960	0.0414	-0.0960	0.0414	-0.0960	0.0414
	γ_{j6}	0	-	0.0639	0.0254	0.1150	0.0325	-0.1371	0.0361
Adjacent category (constant slope)	β_3	-0.1019	0.0138	-0.1019	0.0138	-0.1019	0.0138	-0.1019	0.0138
Adjacent Category (different slopes)	β_{j3}	-0.0762	0.0394	-0.0597	0.0464	-0.1940	0.0651	-0.1187	0.0796
Continuation Ratio (fully constrained)	β_3	-0.1820	0.0224	-0.1820	0.0224	-0.1820	0.0224	-0.1820	0.0224
Continuation Ratio (based on the separate binary logistic regressions)	β_{j3}	-0.1753	0.0328	-0.1835	0.0420	-0.2655	0.0535	-0.1090	0.0801
Continuation Ratio (unconstrained)	β_{j3}	-0.1725	0.0328	-0.1823	0.0421	-0.2577	0.0539	-0.1009	0.0816
Stereotype model (non-linear) using mclest	β_3	0.1594	0.0202	0.1594	0.0202	0.1594	0.0202	0.1594	0.0202
Stereotype model (non-linear) using bootstrapping technique	β_3	-0.1600	0.0316	-0.1600	0.0316	-0.1600	0.0316	-0.1600	0.0316

*** Refer to Table 5.6 for the cut-points fitted for each model**

For the different slopes cumulative model, the 'age-group' and 'sex' covariates have regression parameters over the cut-points, and f-t-education varies by cut-point.
The ϕ_j parameters for the stereotype model (non-linear –estimated using *mclest* (Henrickx (2000)) were ϕ_1=0; ϕ_2=5.0015; ϕ_3=3.3954; ϕ_4=2.1468; ϕ_5=1.
The ϕ_j parameters for the stereotype model (non-linear –estimated using bootstrapping technique) were ϕ_1=0; ϕ_2=5.0831; ϕ_3=3.4511; ϕ_4=2.1790; ϕ_5=1.

Table 7.16: Adjusted Odds Ratios and their 95% confidence intervals for the comparison of less than 13 versus 13 or more years full-time education using the Townsend Disability Scale data

		Cut-point 1*		Cut-point 2*		Cut-point 3*		Cut-point 4*	
Model		***Odds Ratio***	***95% C.I. of Odds Ratio***	***Odds Ratio***	***95% C.I. of Odds Ratio***	***Odds Ratio***	***95% C.I. of Odds Ratio***	***Odds Ratio***	***95% C.I. of Odds Ratio***
Polytomous		0.40	(0.31, 0.52)	0.47	(0.36, 0.61)	0.54	(0.41, 0.70)	0.81	(0.59, 1.11)
Proportional Odds		0.66	(0.58, 0.74)	0.66	(0.58, 0.74)	0.66	(0.58, 0.74)	0.66	(0.58, 0.74)
Unconstrained Partial Proportional Odds	Males	0.76	(0.64, 0.92)	0.67	(0.55, 0.82)	0.48	(0.37, 0.62)	0.41	(0.30, 0.58)
	Females	0.65	(0.55, 0.76)	0.62	(0.53, 0.74)	0.54	(0.45, 0.66)	0.61	(0.47, 0.79)
Adjacent category (constant slope)		0.82	(0.77, 0.86)	0.82	(0.77, 0.86)	0.82	(0.77, 0.86)	0.82	(0.77, 0.86)
Adjacent category (different slopes)		0.86	(0.74,1.00)	0.89	(0.74, 1.06)	0.68	(0.53, 0.88)	0.79	(0.57, 1.08)
Continuation Ratio Model (fully constrained)		0.69	(0.64, 0.76)	0.69	(0.64, 0.76)	0.69	(0.64, 0.76)	0.69	(0.64, 0.76)
Continuation Ratio Model (based on the separate binary regressions/ unconstrained)		0.70	(0.62, 0.80)	0.69	(0.59, 0.82)	0.59	(0.48, 0.72)	0.80	(0.59, 1.10)
Stereotype model		0.45	(0.05, 4.15)	0.58	(0.10, 3.25)	0.71	(0.17, 5.77)	0.95	(0.33, 2.72)

*** Refer to Table 5.6 for the cut-points fitted for each model**

7.4.3 Findings related to the interpretation of the results

Linear Regression Models

The linear models do not adequately summarise the complexity of information given by ordinal data.

- Unlike those findings of Lu (1999) linear regression models provide the same overall conclusion as those of the binary logistic and ordinal regression models – non-smokers are likely to have better overall health compared to smokers and those who have not suffered from a heart attack are likely to have better health than those who have. Also, those with <13 years full-time education are more likely to have less disability compared to those with ≥ 13 years full-time education. However, when relating the predicted mean score back to the ordinal categories, for both datasets all subjects fall into the same predicted response category for the different covariate levels. This would indicate that the analysis based on linear regression models is not sensitive and reliable enough in picking up the complexity of the ordinal data, but may give the investigator a general idea of what is going on.

Binary Logistic Regression Models

Different types of binary logistic regression models (main and interaction) are needed to describe ordinal data in an adequate manner. This, however, causes problems when interpreting the overall results.

- The interpretation of the results from the binary regression models fitted to the Health Status data suggest that the interaction term based on the 1st cut-point was important and could not be ignored. This implies that one cannot use binary models with the same type of effects. This makes it difficult to relate the results to the overall ordinal quality of life scale as different models suggest different relationships of the response and the covariates. Also, one can envisage that this problem would become more apparent as the number of response categories increase.

- The analysis as presented by McGee et al. (MRC CFAS[1] study (1998)) gives results that are based on the disability score (i.e. a continuum) and odds ratios that are based on the allocation of a point, implying unit increase in disability. However, this analysis does not show whereabouts the unit increase occurs (whether its at the beginning or end of the scale). The Townsend Disability Scale on the other hand, is a severity scale and the analysis using ordinal regression models account for the severity levels. Despite the differences in the two methods, both statistical techniques imply that the odds of an increase in disability is more likely for those with less education compared to those with more education. However, the ordinal regression models are more sensitive at picking up changes over the range of the severity of disability.

Polytomous Model

The polytomous model estimates a large number of parameters and fails to account for the ordinal nature of the quality of life scale categories. Also the interpretation of the results is difficult as different odds ratios are used to summarise different cut-points.

- The odds ratios increase at a much greater rate over the cut-points compared to the other models, particularly where proportional odds exists.

- The polytomous model provides cut-point specific odds ratios (and 95% confidence intervals) that monotonically increase for the logits (for both datasets). The drawback of this model is that a large number of odds ratios (i.e. the same number as the logit functions) are used to summarise the data. This makes the interpretation difficult, particularly when the covariates increase. Also, the polytomous model does not take account of the ordering of the categories and any trends in the odds ratios cannot be allowed for. This means that the model is of little use in the context of ordinal data.

Proportional Odds Model

The assumption of proportional odds is stringent and difficult to satisfy. Unless the assumption is satisfied, the odds ratios cannot be used to summarise the data, as this will lead to misleading interpretations.

- For both datasets, the violation of the assumption of proportional odds was apparent when the model was fitted.

Partial Proportional Odds Models

The unconstrained partial proportional odds model is an efficient ordinal regression model in that it allows some covariates to have proportional odds and other covariates to vary by cut-point. This simplifies the interpretation of the odds ratios.

The constrained partial proportional odds model is of little use, as the constraint parameters for this model are determined using the odds ratios obtained from the unconstrained model and therefore there is little difference in the estimates of the odds ratios.

- The unconstrained partial proportional odds model is useful in summarising the data as it simplifies the interpretation of the odds ratios. The unconstrained partial proportional odds model is 'half-way' between the polytomous and the proportional odds model in the number of parameters it fits. It achieves model parsimony without indication of lack-of-fit.

- The unconstrained and constrained models provided very similar estimates of the odds ratios and their 95% confidence intervals. Given several covariates in the model with more than one covariate where non-proportionality exists, the constraints are difficult to determine for the constrained partial proportional odds model (Townsend disability score). Furthermore, using the odds ratios from the unconstrained version of the model to derive the constraints for the constrained model can be problematic, as one is observing the data to obtain the constraints and then a model is fitted using these (Health Status data).

- The 1st order interaction terms of the unconstrained partial proportional odds model illustrated that when proportionality exists for the two covariates the interpretation of the odds ratios is much simpler than if one of the covariates has non-proportionality.

Adjacent Category Model

(a) Constant Slope Model

The assumption of the constant odds model is stringent and is difficult to satisfy. However if the assumption holds, then the constant slope adjacent category model provides by far the simplest interpretation, as there is only one common odd ratio.

- There are two factors in favour of this model: (i) this model allows for the ordering of the categories; (ii) it provides model parsimony. Whether or not one uses this model to summarise the data, is largely dependent on the choice of the modelling strategy. It may be that all the covariates are of interest, and the assumption has to be equally satisfied equally by them. The chance of this happening is quite remote (as in the Health Status data) and therefore it is likely that the adjacent category model will not be a chosen model. If one is interested in a single covariate, then the constant slope assumption has to be satisfied by this variable and the adjusted covariates would need to be 'forced' to satisfy this assumption (if this was not the case).

(b) Different Slopes Model

The different slopes adjacent category model fits the same number of odds ratios as the polytomous model, but allows for the ordering of the categories (through the formation of the logits). This would imply that the model was more efficient than the polytomous model, but not as efficient as the unconstrained partial proportional odds model. Also the interpretation of the odds ratios is difficult, as there are as many odds ratios as cut-points.

- The odds ratios of the different slopes adjacent category models for both dataset were found to fluctuate over the cut-points. This would imply that without estimating the odds ratios it is difficult to know how they are likely to behave over the cut-points.

Continuation Ratio Model

The assumptions of the fully constrained continuation ratio model are again difficult to satisfy.

The different slopes model is best fitted using the unconstrained model as opposed to the separate binary logistic regressions.

The unconstrained continuation ratio model has the same number of odds ratios as the different slopes adjacent category model and also allows for the ordering of the response categories through the formation of the logits. Therefore in terms of efficiency it is very similar to the different slopes adjacent category model.

There is a possibility of reducing the odds ratios by constraining them to be equal, and the resultant is the partially constrained continuation ratio model. This model is as efficient as the unconstrained partial proportional odds model.

- The binary logistic regression models have to be fitted with redundant parameters, so as to allow for consistency in the type of parameters used (as in the Health Status data). The interaction term of the binary logistic model based on the first cut-point cannot be ignored, as it produces results that are quite important. However, it has to be borne in mind that the binary logistic regression based on the continuation ratio logits do not allow for the ordinal categories in the same way as the unconstrained continuation ratio model. The contribution of this interaction effect is negligible in the overall interaction term when fitting the unconstrained continuation ratio model, where the ordering of the categories is taken into account (in one model). This would indicate that although there is some evidence of a difference between the 'heart attack' categories in the smokers and non-smokers group, given the '*good*' and '*excellent*' health status categories only, when one takes account of the entire ordinal health status scale, this difference is eliminated. On the basis of this, it would appear that the unconstrained continuation ratio model was a more efficient model compared to the binary model.

- The binary and unconstrained models provide identical parameters and odds ratios, when there are the same number and type of parameters for both models. The unconstrained continuation ratio model fitted the same number of odds ratios as the polytomous and the different slopes adjacent category model. The odds ratios for the unconstrained model suggested some homogeneity. The partially constrained model was fitted without sacrificing model fit and at the same time increasing model parsimony.

Stereotype Model

The stereotype model is only ordinal if the ordinal parameters (ϕ_j) are ordered. Furthermore the estimates of the odds ratios are not comparable to the other ordinal regression models.

The stereotype model provided ordering parameters that satisfied the inequality (3.26), suggesting that this model was an ordinal regression model. However, these parameters were not monotonic for the stereotype model fitted using the Townsend Disability Scale data, suggesting that this model did not adequately allow for the ordinality in the *y*-response with respect to the covariates. The odds ratios of the stereotype model (as for the polytomous model) increase over the cut-points at a greater rate compared to those of the other models, as illustrated by the data from the health status categories. In fact the odds ratios based on the final cut-point for the 'no heart attack' and 'heart attack' comparison are almost twice the odds ratios for the other models. These odds ratios are only comparable to the polytomous model and not to any other ordinal regression model.

The odds ratios provide further evidence that a 2-dimensional stereotype model is not required to summarise the quality of life data.

The odds ratios of the stereotype model increase over the response categories for both covariates, given the Health Status data. This would suggest that the both predictors behave in a similar fashion over the categories and that we do not have a case of different predictors explaining different parts of the ordinal scale.

7.5 Comparison of the models

7.5.1 Parameter estimates and the odds ratios

There was a violation of the constant slope assumption for proportional odds, adjacent category and the continuation ratio models (Health Status data). However, the adjacent category and the continuation ratio models fitted for Townsend Disability Scale data were assumed to satisfy the assumption. Although the comparison below has been made with the constant and different slopes models, one should be cautious due to these violations.

For both datasets, the cut-points used for the cumulative logit models and the binary logistic regression models were the same. However, the estimates of the parameters (standard errors) and odds ratios (with 95% confidence intervals) differed for these models. For the Health Status data, the parameters and odds ratios for the binary logistic regression models were cut-point specific (as for the different slopes cumulative model), for the proportional odds they were constant over the cut-points and for the partial proportional odds models it was the combination of the two. For the Townsend Disability Score dataset, the binary and proportional odds models assumed constant odds, whereas the partial proportional odds models assumed otherwise.

The polytomous model, different slopes adjacent category and unconstrained continuation ratio models all shared very similar estimates/odds ratios for the first cut-point. Also the unconstrained continuation ratio, different slopes cumulative logit and the binary logistic regression models all had the same parameter estimates for the final cut-point.

(a) Binary Logistic Regression Models

Health Status

The confidence intervals of the main effects binary logistic regression model were wider than those of the proportional odds model, where proportionality was present (i.e. the 'heart attack' covariate) (also found by Scott et al. (1997)). Also as stated by Armstrong and Sloan (1989), at the optimal point (i.e. for cut-point 2) the odds ratios from the main effect binary logistic model were quite similar to the main effects proportional odds model provided proportionality was present (Table 7.10). However if proportional odds was not found to be satisfied (as in the 'smoke' covariate), then these findings did not hold, even when the proportional odds model was fitted (with the assumption of proportionality imposed on the covariate on the model).

Manor et al. (2000) found similarity in the results when using the logistic and ordinal regression models. Likewise, a similarity in the results was noted when comparing the main effects binary logistic regression models with the main effects cumulative ordinal regression models.

Townsend Disability Scale data

The adjusted odds ratio (and 95% confidence interval) of the binary model given in Table 7.14 is very similar to the ones produced by the proportional odds and the fully constrained continuation ratio models in Table 7.16. However, the model assumption was violated in the proportional odds model and it was only satisfied by 'full-time education' in the fully constrained continuation ratio model. Thus the results would suggest that if the cut-point specific odds ratios can be assumed to be homogenous, then the binary logistic model as fitted by McGee et al. (MRC CFAS[1] study (1998)) would be a satisfactory way to analyse the data. If the assumptions are violated, then the results are misleading, as a single summary measure (provided by the binary model) would be used to describe data that may otherwise provide the odds ratios which would be varying, at worst, quite substantially over the cut-points. This highlights the fact that ordinal regression models are much more sensitive at picking up the varying information over the response categories, which is thought to be constant when fitting the binary model as presented by McGee et al. (MRC CFAS[1] study (1998))

(b) Cumulative Odds Models

The odds ratios of 'smoke' in Table 7.9 for the different slopes cumulative odds model and the two partial proportional odds models are the same for each of the cut-points. The unconstrained partial proportional odds model is effectively fitting β_j (for a given covariate) and this parameter has been separated into two components β and γ_j. The reason this model is fitted in this way is because the assumption of proportional odds is easier to test (see section 3.5.3). Also the comparison of these models with the proportional odds model provides an insight into where the difference lies and how the violation of the proportional odds has come about (for the Health Status data the odds ratio for final cut-point of 'smoke' was substantially different for the constant and different slopes models).

For both datasets, the estimates of the unconstrained partial proportional odds model are monotonically increasing or decreasing. Bender and Grouven (1998) state (see Chapter 3: section 3.11) that the partial proportional odds model is more efficient than the polytomous and binary logistic regression models, and this was certainly found to be the case from the data used in this thesis.

The proportional odds model produces estimates with smaller standard errors compared to the unconstrained partial proportional odds model and the different cumulative odds models (in particular for the 'smoke' covariate). This is because one is providing a single parameter estimate using increased information. However, in the different slopes models, one allows for the cut-points and the estimated parameters are based on these cut-points (with a reduced number of data which are cut-point specific).

(c) Adjacent Category Models

The parameter estimates and odds ratios of the different slopes adjacent category model fluctuate over the cut-points for both datasets (see Table 7.6, 7.9 and 7.10). The 95% confidence intervals of the constant slope model are narrower than the different slopes model. This is due to the same reasons stated in (b).

Ananth et al. (1997) demonstrated similarity in the odds ratios provided by the different slopes adjacent category model and the polytomous model (see Chapter 3: section 3.11). This was not found to be the case for both datasets presented in this thesis. It was difficult to examine why this was the case, as the data quoted in the paper by Ananth et al. (1997) was found to be incorrect (Cole, 1999).

(d) The Continuation Ratio Models

The proportional odds model has been much compared to the constant slope continuation ratio model in the literature (see Chapter 3: section 3.11). The similarity in the results was apparent from Tables 7.9, 7.10 and 7.16. The parameter estimates that were the closest to the proportional odds model were those provided by the constant slope continuation ratio model.

The parameter estimates for the different slopes continuation ratio model (based on the binary logistic regressions/unconstrained continuation ratio model) fluctuated over the cut-points. For the unconstrained model and the binary logistic models used to fit the continuation ratio logits, provided the type of effects fitted are the same in both models then the parameters estimates are identical (see Table 7.16). The 95% confidence intervals for the constant slope model were narrower than those of the different slopes model.

(e) The Polytomous and the Stereotype Models

The odds ratios were found to be monotonically increasing when fitting the polytomous models to the two datasets. It can be shown that if proportional odds exist for a given model then when the polytomous model is fitted the odds ratios will always be monotonically increasing (see Appendix IV).

Like Cox and Chaung (1984) and Manor et al. (2000), we found that the polytomous model was a better fit than the other ordinal regression models.

We found that the stereotype model was of little value for the two quality of life measures considered. This was unlike the findings of Greenland (1994), who valued the stereotype model. However, in his examples only one covariate with an ordinal outcome was used to fit the models. Furthermore, the need for a two-dimensional stereotype model did not arise.

The odds ratio of the polytomous and stereotype model are substantially different to the odds ratios of other models (see Tables 7.9, 7.10 and 7.16). For both models the odds ratios increase monotonically. This would suggest that despite the constraint parameters estimated in the stereotype model, there is not much difference in the odds ratios given the two models.

7.5.2 Comparison of the models using statistical inference

(a) Comparison of the different cumulative logit models

The test of whether the proportional odds model was as good a fit as the unconstrained partial proportional odds model was constructed using the Health Status data. It was found that the unconstrained partial proportional odds model was a better fit than the proportional odds model (H_0: $\gamma_{21} = \gamma_{31} = 0$: $\chi^2_2 = 15.62$; $p = 0.01$). Furthermore, this was a more parsimonious model than the one that allowed for separate slopes for the cut-points for each covariate (since 7 parameters were estimated in the unconstrained partial proportional odds model and 9 parameters were estimated in the different slopes model).

We found that the constrained partial proportional odds model was as good a fit as the unconstrained partial proportional odds model (H_0: $\gamma_{31} = 40\gamma_{21}$: $\chi_1^2 = 0.00$; p=0.997).

The comparison of the unconstrained partial proportional odds model and the proportional odds model was a test of proportional odds for the 'full-time education' covariate, for the Townsend Disability Scale data. This is given in section 7.2.2.1 (b).

(b) Comparison of the constant slope and different slopes adjacent category models

The comparison of the constant and different slopes adjacent category models provided a means of testing the constant slope assumption and this is discussed above (sections 7.2.1.1 and 7.2.2.1).

(c) Comparison of the constant slope and different slopes continuation ratio models

Again the comparison of these two types of models provided a means to test the assumption of constant slope over the cut-points and this is as discussed above (sections 7.2.1.1 and 7.2.2.1).

(d) Comparison of the models using the Akaike Information Criteria (*AIC*)

The Akaike Information Criteria (*AIC*) was obtained for the polytomous, proportional odds and the stereotype models. It could not be derived for the other models, as the analysis for these models did not provide a log-likelihood value (the method of estimation was weighted least squares).

The *AIC* values are displayed in Table 7.17. This statistic allows one to make a comparison of ordinal models having adjusted for the number of parameters they fit. Generally there is indication that the polytomous and stereotype models provide similar fit to the two datasets. Also for the Health Status data, there was not much difference in the *AIC* statistic for the proportional odds and the stereotype model.

Table 7.17: Akaike Information Criteria (AIC) statistics

Model	***AIC* statistics**	
	Health Status data	***Townsend Disability Scale data***
Polytomous	29347.16	34344.19
Proportional odds	29353.84	30675.55
Stereotype	29353.73	34470.41

7.5.3 Findings of the comparison of models

The following were found when comparing the different regression models.

- *The binary logistic models give similar estimates of odds ratios to the cumulative logits models, although the 95% confidence intervals are wider. Therefore provided the same effects are used for the binary and cumulative models then the binary analysis allows one to summarise results that could otherwise be obtained using the cumulative logits models. Also, the binary model fitted by McGee et al. ((MRC CFAS[1] study (1998)) provides similar results to the constant slope ordinal regression models.*

- *The odds ratios of the proportional odds and fully constrained continuation ratio models are very similar.*

- *The unconstrained partial proportional odds model is identical to the different cumulative slopes model.*

- *Constant slope models provide more precise estimates of odds ratios compared to models where cut-point specific odds ratios are given.*

- *The Akaike Information Criteria (AIC) statistics indicated that the fit of the proportional odds and the stereotype models were similar, having adjusted for the number of parameters fitted. This would suggest that the stereotype model is as efficient as the proportional odds model. However, the AIC could not be computed for all models. There is therefore scope for further research.*

- *The odds ratios of the stereotype and polytomous models are very similar. This suggests that the ordering constraints in the stereotype model have done little to improve on the polytomous model. In fact, where the polytomous model provides a good fit model, when the ordering parameters are imposed, the model becomes too constrained leading to a lack-of-fit model (as in the Health Status data). Once again, unlike evidence provided by some authors, there was little evidence in this thesis to suggest that the stereotype model is an ideal model for analysing discrete ordinal response data.*

7.6 The 'best fit' models

In the light of all the evidence provided from the results of the analysis, given the Health Status ordinal scale, the 'best fit' models were:

- the unconstrained Partial Proportional Odds model;
- the partially constrained Continuation Ratio model.

These models provide parsimony as well as adequate fit to the data.

Given the Townsend Disability Scale data, the 'best fit' models were:

- the fully constrained Continuation Ratio model;
- the constant slope Adjacent Category model.

These models provide parsimony. However, although they do not fit the data well, the models still provide an informative summary. The value of goodness-of-fit procedure was to assess whether the model fits and if this is not the case, where the lack of fit lies. Its purpose was not to formally reject the model.

7.7 Summary

- ***Linear Models***: Linear models are not adequate for analysing ordinal quality of life data. The evidence that supports this statement is provided by both datasets in this thesis. In brief (i) the assumptions of the model could not be checked due to the nature of the data; (ii) the standard errors of the parameter estimates indicated a constant variance over the y-response and this was unlikely due to the nature of the data; (iii) despite the lack-of-fit, the Normal plots were unable to highlight the outliers or influential observations, again due to the ordinal nature of the data, and (iv) the results and their interpretation did not adequately summarise the complexity of the ordinal scale.

- ***Binary Logistic Regression Models***: Binary logistic regression models (separate models and the model fitted by McGee et al. ((MRC CFAS[1] study (1998))) are only adequate for analysing ordinal data, if one can assume constant odds over the ordinal scale. If one cannot assume constant odds, then this method will provide an incorrect conclusion, as there is implication then that the odds vary over the range of the scale. Fitting separate binary models would allow one to model the varying odds ratios. However, this may still present problems as different cut-point specific binary models may be used and this would make the interpretation of the overall ordinal scale difficult (as was shown using the Health Status data). Also, there is a danger of spuriously significant effects emerging due to multiple models fitted over the scale.

- ***Ordinal Regression Models***: All ordinal regression models are as efficient as the binary logistic model in terms of the Asymptotic Relative Precision, as this measure is not affected by the type of parameters fitted, but rather by the number of covariates included.

- ***Constant Slope Ordinal Regression Models***: The constant slope ordinal regression models (namely the proportional odds, adjacent category and the continuation ratio models) all have the same number of fitted parameters. The constant slope ordinal regression models are, in theory, the best models for analysing ordinal data as they provide a single summary measure and are therefore parsimonious. If the assumptions hold, then these models are by far the simplest to interpret and provide more accurate point estimates. However, in practice, the model assumptions that support this single measure are too stringent, as was demonstrated by the Health Status data. This will also produce a lack-of-fit model and the model cannot be used to report the results. It may be that the model assumption is only satisfied by the covariate of interest. Then the other covariates can be 'forced' to satisfy the assumptions (as in the Townsend Disability Scale data), resulting in a model that can be used to summarise the results but lacks goodness-of-fit. Also if the assumptions hold, the binary logistic models may be an adequate method for analysing the ordinal data.

- The global χ^2-score test statistic is of use when testing the constant slope assumption for all covariates in the model. It is of little use when one is interesting in a covariate where the assumption is required to hold, and if this is the case, then the other covariates are 'forced' to satisfy the model assumption.

- ***Partially Constrained Models***: Given that the constant slope assumption does not hold, in practice, the partially constrained models (the unconstrained partial proportional odds and the partially constrained continuation ratio models) are more appropriate for fitting ordinal data. These models do not fit as many parameters as the different slopes models and the parameters can be constrained to be equal (either all or some of them), without sacrificing the goodness-of-fit. The interpretation of the odds ratios is simpler than that provided for the different slopes models.

- ***Different Slopes Regression Models***: The different slopes regression models (namely the polytomous, different slopes cumulative, different slopes adjacent category and unconstrained continuation ratio models) all have the same number of estimated parameters. The polytomous model does not account of the ordering of the y-response and therefore is not considered as an ordinal regression model, but rather a model used to examine multinomial data. These models make no assumptions regarding the parameters, and therefore a parameter is fitted for each cut-point and covariate in the model. This provides a good fit model, due to the large number of odds ratios, However, the interpretation is somewhat difficult and cumbersome.

- ***Best Fit Models***:

From the health status data, the best fit models were:

  - Polytomous model;
  - unconstrained Partial Proportional Odds model;

- Partially constrained Continuation Ratio model;
- Different slope Adjacent Category Model;
- Different slope Continuation Ratio Model.

From the Townsend Disability data, the most appropriate models that could be used to describe the data were:

- Polytomous model;
- Unconstrained Partial Proportional odds model;
- Constant slopes Adjacent Category model;
- Constant slopes Continuation Ratio model;
- Different slope Adjacent Category model;
- Different slope Continuation Ratio model;
- Stereotype model.

- ***Stereotype Model***: Greenland (1994) strongly argues in favour of the stereotype model. There was little evidence in this thesis to suggest that the stereotype model was of significant use for analysing ordinal data. The main areas identified as problematic were:

(a) The stereotype model was initially devised using the polytomous model. It was devised to (i) allow for the ordinality of the *y*-response (which the polytomous model fails to do) and (ii) reduce the number of parameters estimated in the polytomous model, in the aim of achieving model parsimony. In practice, it is likely that the parameters that are estimated will not necessarily be ordered (as in the Townsend Disability Scale data) and the constraints imposed that allow for model parsimony are too stringent, leading to a poor fitting model.

(b) *Dimensionality*: Only the one dimension stereotype model was found to be of use in fitting the ordinal data; the two dimensional model was of limited use.

(c) *Indistinguishability*: Models with and without the indistinguishability criteria could not be compared due to lack of compatibility of statistical software.

(d) *Ordering*: There is a need to compute the 95% confidence intervals of the ordering parameters (ϕ_j), in order to assess the precision of the point estimates. This has not been cited in the literature.

(e) *Bootstrapping*: This technique is computationally intensive and required a very lengthy computer run-time.

(f) *Odds ratios*: The interpretation of the odds ratios is very similar to the polytomous model, implying that in terms of the end product, the stereotype model has done little to improve on the polytomous model.

(g) *Goodness-of-fit*: The linear model provides a way of assessing the goodness-of-fit of the non-linear model.

- ***Model Comparison***: The comparison of the models suggested:

(a) The unconstrained continuation ratio model was a more efficient method for fitting the continuation ratio logits compared to the separate binary models.

(b) The goodness-of-fit statistics of the unconstrained continuation ratio model should be identical to the sum of the chi-squared values obtained from the individual logistic regression models. However, it was not possible to check this as different procedures were used to compute the models.

(c) The odds ratios produced by the unconstrained partial proportional odds, polytomous and the stereotype models were found to always increase (or decrease) monotonically. The odds ratios of the polytomous and the stereotype models increase at a greater rate over the cut-points compared to the odds ratios of other models. The odds ratios produced by the different slopes adjacent category or the unconstrained continuation ratio models were found to fluctuate over the cut-points.

(d) The linear version of the stereotype model could be used to check the goodness-of-fit of the non-linear model.

- ***Sparse data***: Presented with a skewed data distribution on a scale and sparse data in a category at one of the extreme ends of the scale, then poor fit results when this category is used as a single data point in the formation of the logits and the difference between the marginal probabilities (one of which is the category where there are sparse data) that form these logits is relatively large.

- ***Interaction terms***: The interpretation of an interaction term in an ordinal regression model varies depending on whether the constant slope assumption holds for the cut-points of the main effects that constitute the interaction term.

- ***Goodness-of-fit***: The Lipsitz's (1996) procedure for testing the goodness-of-fit of the models was of little use. Instead the Likelihood Ratio test (in the case of models fitted using maximum likelihood estimation) and the Wald goodness-of-fit statistic (for models fitted using weighted least squares) were used.

- ***Residual Analysis:*** Residual analysis should be carried out using the separate binary analyses. However this is not always possible, and therefore one has to rely on assessing the observed and fitted logits/probabilities. The residuals of the cell probabilities were found to be of little use. The residuals of the logits were more sensitive at picking up the outliers/influential observations. However, this can only be based on visual assessment.

CHAPTER 8 - DISCUSSION AND CONCLUSION

The primary purpose of this chapter is to illustrate the most appropriate ordinal regression models given the two instruments used to assess aspects of quality of life in elderly people. In order to achieve this aim, I begin the discussion with section 8.1 –this outlines further aspects related to ordinal regression models that have emerged from the results. Then drawing the results and findings together from both Chapter 7 and section 8.1, the 'best' fit models are detailed, given the two quality of life measures used in this thesis. Finally, in section 8.2 this thesis is concluded by reflecting on the hypotheses initially outlined in Chapter 1. Also, the conclusion highlights the contribution of this study to the literature, its limitations and recommendations for further work.

8.1 Discussion

8.1.1 Assumptions of constant slope for the models fitted using the Townsend Disability Scale data

The likelihood of all the covariates satisfying the assumption of constant slope is quite small, given a large number of covariates. As a result, a pragmatic view has to be taken, and this was certainly the case when fitting the covariates in the ordinal regression models using the Townsend Disability Scale data. The strategy adopted here for including the covariates in the model was similar to that used in survival analysis (Parmar and Machin, 1995). In the latter, provided the covariate of interest satisfied the proportional hazard assumption then the adjusting covariates are assumed to satisfy the assumption (even though this may not necessarily be the case).

8.1.2 Partially constrained Adjacent Category Model

The partially constrained version of the adjacent category model is not quoted in the literature. However, using the different slopes adjacent category model for the Health Status data, the hypotheses H_{01}: $\beta_{11}=\beta_{21}$ and $\beta_{21}=\beta_{31}$ (based on Table 5.1 and equation (3.17)) for the 'smoke' covariate and H_{01}: $\beta_{12}=\beta_{22}$ and $\beta_{22}=\beta_{32}$ for the 'heart attack' covariate were constructed and tested. This was done in a very similar way to the partially constrained continuation ratio model. There was indication that $\beta_{11}=\beta_{21}$ ('smoke': χ_1^2 =0.46 with a p

=0.49) and β_{12}=β_{22} ('heart attack': χ_1^2 =2.23 with a p =0.14). Also there was not enough evidence to suggest homogeneity for the parameters β_{21} and β_{31} ('smoke': χ_1^2 = 10.18 with a p =0.001) and for β_{22} and β_{32} ('heart attack': χ_1^2 = 11.22 with a p =0.001). The partially constrained adjacent category model was fitted such that the log odds for the first two cut-points (i.e. '*good*' v. '*excellent*' and '*fair*' v. '*good*') were constrained to be the same for both covariates. The parameter estimates (standard errors) and the odds ratios (and their 95% confidence intervals) of this latter model are displayed in Table 8.1. The Wald test suggested the goodness-of-fit was adequate (χ_5^2 =9.37 with a p =0.10). The odds of '*good*' health were equal to that of '*excellent*' health for those who smoked (compared to the non-smokers). Also the odds of having '*fair*' health were identical to that of '*good*' health for the same group of subjects. However, the odds of '*poor*' health were 1.5 times that of '*fair*' health for the smokers compared to the non-smokers.

From Table 8.1 the odds of '*good*' are twice that of '*excellent*' health. This is the same as the odds of '*fair*' health as opposed to '*good*' health for those who suffered from a heart attack as opposed to not suffering from one. Also, the odds of '*poor*' health are 1.45 times that of '*fair*' health for those who have suffered from a heart attack (compared to those who have not suffered from one).

Table 8.1: Parameter estimates (with standard errors) and the odds ratios (with 95% confidence intervals) from the Partially constrained Adjacent Category model

covariates	**'good' v. 'excellent'**	**'fair' v. 'good'**	**'poor' v. 'fair'**
	Parameter estimates (standard errors)		
Smoke	0.03 (0.02)	0.03 (0.02)	0.21 (0.05)
Heart attack	0.37 (0.02)	0.37 (0.02)	0.19 (0.05)
	Odds ratios (95% confidence intervals)		
Smokers v. non-smokers	1.05 (0.98, 1.12)	1.05 (0.98, 1.12)	1.51 (1.24, 1.82)
Heart attack v. no heart attack	2.11 (1.93, 2.32)	2.11 (1.93, 2.32)	1.45 (1.19, 1.77)

Like the unconstrained partial proportional odds model and the partially constrained continuation ratio model, this model proved to be parsimonious as well as providing a good-fit.

These results further emphasise the evidence found in Chapter 7 – in pragmatic terms, where the chances of constant odds is reduced, the partially constrained models are most appropriate for analysing ordinal outcomes.

8.1.3 The Partial Proportional Odds Models

The constrained partial proportional odds model is a simplification of the unconstrained partial proportional odds model in terms of the parameter estimates. However, it is problematic, as already mentioned, in that the constraints are estimated using observed data. In theory, one may be able to use the methodology of the stereotype model to estimate the constraint parameters.

Peterson and Harrell (1989, 1990) fit the unconstrained partial proportional odds model using the equation (3.14). The proportionality assumption is easier to test when the model is specified using this format. The model can also be fitted as:

$$\log\left[\frac{\Pr(Y_i \leq y_j)}{\Pr(Y_i > y_j)}\right] = \alpha_j + \sum_{k=q+1}^{p} x_{ik}\beta_k + \sum_{k=1}^{q} T_{ik}\beta^{*}_{jk} \qquad j = 1 \ldots c\text{-}1 \qquad (8.1)$$

where there are $p - q$ covariates with proportional odds and q covariates which have non-proportionality. For these latter covariates, separate regression parameters are fitted over the cut-points. The test of proportionality would be $\beta^{*}_{2k} = \beta^{*}_{3k} \ldots\ldots = \beta^{*}_{(c-1)k}$ for each of the q covariates. The constrained partial proportional odds model could then be fitted as:-

$$\log\left[\frac{\Pr(Y_i \leq y_j)}{\Pr(Y_i > y_j)}\right] = \alpha_j + \sum_{k=q+1}^{p} x_{ik}\beta_k + \sum_{k=1}^{q} T_{ik}\Gamma_j\beta^{*}_{k} \qquad j = 1 \ldots. c\text{-}1. \qquad (8.2)$$

The non-proportional component of the R.H.S of equation (8.2) is similar to the ordering constraints/beta component of the R.H.S of equation (3.23). This would suggest that the constraint parameters, Γ_j in equation (8.2), rather than estimated using the unconstrained partial proportional odds model, could be estimated using a non-linear version of the constrained partial proportional odds model. However there are some fundamental differences between models (8.2) and (3.23). For instance, the constraints in equation (3.23) apply to all the covariates, where as in equation (8.2) they only apply to covariates with non-proportionality. As a result, Hendrickx's (2000) macros cannot be easily used at present to estimate the constraints in equation (8.2) and therefore further research is required to adapt these macros for the partial proportional odds models. Also, the purpose of the constraints are very different for both equations: for the stereotype model these serve to order the y-response with respect to the covariates so as that the relationship can be described as ordinal, and the same constraints have to be shared by all the covariates. For the constrained partial proportional odds model, the purpose of the constraints is to simplify the interpretation of the odds ratios over the cut-points, and it is not necessary that the same constraints be used for all covariates. Therefore if the constraints fail to simplify the interpretation, there is little purpose in estimating them.

8.1.4 Different slopes Cumulative Logit Model

The conclusions from the binary logistic regression models using the Health Status data (in particular the model based on the 1st cut-point), suggested some unusual results based on the interaction term (see section 7.4.1.2). To assess this further, the different slopes cumulative logit model was re-fitted. Up to now, the purpose of this model was purely to visually examine the cut-point specific odds ratios in the view to fit/not fit a proportional odds model. However, the different slopes cumulative model (as fitted using the Health Status data) is based on the same logits as the binary models. In theory, both models should provide very similar results (if a difference occurs, it is due to the fact that the parameter estimates of the binary logistic models are based on maximum likelihood estimation and the parameters of the different slopes model are based on the weighted least squares method). This was indeed the case when examining the odds ratios given by the main effect models (binary and different slopes cumulative) in Tables 7.8 and 7.9. This provided a basis for comparing the results of the two models.

The interaction term of 'smoke' and 'heart attack' was fitted in the main effect different slopes cumulative model and this term was non-significant (χ^2= 6.56 on 3-df.; p= 0.09). This suggested that the interaction effect of the binary logistic regression model based on the first cut-point although significant on its own had little contribution when the entire ordinal scale was taken into account. These results were similar to those produced for the binary models (based on the continuation ratio logits) and the unconstrained continuation ratio model (see section 7.4.1.6) and further highlighted the concerns of using binary logistic models for ordinal data.

8.1.5 Drawing the results and findings together with emphasis on health-related quality of life instruments used in elderly people

Data on health-related quality of life are often captured on instruments that are either single- or multi-item (section 1.1.2). In this thesis, the Health Status scale is a single global measure for overall health, whereas the Townsend disability score is a multi-item measure for disability. Also single-item scales often have ordinal categories that are discrete in nature and are termed as 'assessed' variables, whereas multi-item ordinal scales measure some underlying continuum and are termed as 'group continuous' variables (section 1.1.2).

The assumptions (see section 3.5.2 (b)) of the proportional odds (and partial proportional odds) models are ideally suited for analysing grouped continuous variables. These models allow for the fact that the total score of a group of items is a reflection of the response provided by a subject for the dimension of quality of life being assessed. Then when modelling using the cumulative odds models, the main interest is not in the response categories, but rather in the total score and its distribution and the information about the score obtained through the ordinal categories. If the cut-points were removed or more added in, then, the results would not change, due to the latent underlying variable. In the literature, no mention is made of whether other ordinal regression models have the capacity of accommodating for the total score in the same way as the cumulative odds models. However, one would expect that the Continuation Ratio model (where the categories are accumulated in the denominator of the logits) would have to allow for the assumption of an underlying continuum. The polytomous, adjacent category and the stereotype models do not account for the total continuous score of the items directly, but the score is used in the formation of the ordinal categories (i.e. 1='*excellent*', 2='*good*', ...). The implication here is that the cumulative logit and continuation ratio models are more ideal for analysing grouped continuous outcomes, since they incorporate the underlying quality of life score into the analysis.

The single global item (such as the Health Status score), on the other hand, assumes that the ordinal categories are discrete and there is no overall numerical measure (e.g. a total score) from this scale that reflects the aspect of quality of life considered. Instead, for a given subject individual labelled categories on the scale (e.g. '*excellent*', '*good*', '*fair*' or '*poor*') is taken as the 'total score'. According to McCullagh (1980), it is not necessary to suppose the existence of an underlying continuous variable in order to use the proportional odds/ partial proportional odds models (see section 3.5.2(b)). He argues that this model can be adapted for 'assessed' ordinal data, but the interpretation of the parameters will suffer. In the context of quality of life, accumulating the ordinal categories is meaningless, and at worst, the accumulated categories take on a changed form. For instance, for the Health Status score, accumulating categories ('*fair*', '*poor*') versus ('*excellent*', '*good*') implies 'fair or less than fair' versus 'better than fair' health status. If the latter were presented on an independent binary scale, the rated response would possibly be different to that of the former grouped ordered categories. This situation is analogous to that found in questionnaire designs where the change in form of words of a question corresponding to the amalgamation of two categories will not necessarily result in the amalgamation of responses in the two categories.

With this in mind, despite McCullagh's (1980) argument, the proportional odds/ partial proportional odds models are less attractive for analysing assessed quality of life measures. Likewise, the continuation ratio models, where the logits are based on cumulated cell probabilities are also less appropriate for the latter situations. The polytomous, adjacent category and the stereotype models are based on logits formed using individual discrete response categories and therefore are more attractive for assessed ordinal scales.

From this we deduce:

- *for 'group-continuous' scales, the proportional odds, unconstrained partial proportional odds and the continuation ratio models are more attractive.*

- *For 'assessed' response scales, the adjacent category, stereotype and the polytomous models are more appropriate.*

8.1.6 The most appropriate ordinal regression models given quality of life instruments

In this thesis, the 'best fit' models given the two health-related quality of life measures are stated in section 7.6. In addition to these, it was found that the partially constrained adjacent category model was also a 'best fit' model given a discrete ordinal response scale (see section 8.1.2).

In the light of the above comments and the findings in the results, the following is stated:

- *Given a single-item assessed quality of life measure, the most appropriate model is the partially constrained Adjacent Category model.*

- *Given a multi-item quality of life measure, the results and findings suggest that the most appropriate model is the fully constrained Continuation Ratio model.*

The partially constrained Adjacent Category model is not mentioned in the literature. It was derived in the light of other partially constrained models. The reasons for the choice of this model are (i) it allows for parsimony; (ii) it does not sacrifice the goodness-of-fit and (iii) it accommodates for the discrete nature of the ordinal health status categories. It is the only model from the partially constrained set (unconstrained partial proportional odds, partially constrained continuation ratio and the partially constrained adjacent category) to fulfil these necessary requirements of a 'good fit' model, given a discrete response quality of life scale. These results can be easily generalised to other discrete ordinal scales.

The results of the Townsend Disability Scale data are somewhat less generalised. The reason of this is that the adjusted covariates were 'forced' to satisfy the constant slope assumption and with this in mind, the covariate of interest also had constant slopes. This is not necessarily a result that one would expect each and every time, given several covariates. There is a possibility that despite the model assumption being 'forced' upon the adjusting covariates, it is not satisfied by the covariate of interest. There is therefore some suggestion here that given a group continuous quality of life scale, the unconstrained partial proportional odds or partially constrained continuation ratio models may be the more likely models used to generalise the results. These models take account of the group continuous nature of the ordinal scale and are also likely to offer model parsimony without sacrificing the goodness-of-fit.

8.1.7 The quality of life instrument and the study population

The choice of the model is largely dependent on the type of instrument used (whether assessed or group continuous). However, many quality of life instruments are generic, such as the Health Status index from the SF-36 questionnaire and are used widely on different types of subjects. It is possible to choose an ordinal regression model based on the type of instrument used and type of study population considered.

Some ordinal regression models have been devised with specific types of study populations in mind. For instance, the logits of the continuation ratio models would best describe data on very old and frail respondents or even elderly people with certain types of cancer. This is because the response on these subjects is likely to deteriorate or at the best remain constant for some period of time (and this is best captured using continuation ratio logits). Alternatively, one may be studying the quality of life on a general group of the elderly population and the objective is to assess the response as captured on the scale. The 'best' models under these circumstances would be the cumulative logit models. For a group continuous quality of life scale, one of the latter two ordinal regression models can be applied depending on the study sample. However, if one was to assess the health of very sick and frail elderly subjects using, for example, the Health Status index, then although the continuation ratio logits may provide the most appropriate odds ratios, one would have to bear in mind the difficulties of interpreting the results since a discrete ordinal scale was administered. Alternatively, if one is only interested in comparing the 'referent' category response with the other response categories and a multi-item instrument scale was administered, then the drawbacks of the models that compute the generalised logits, given data based on the total score, would have to be taken into account.

8.1.8 Self-rated health

As mentioned in Chapter 3, in the literature, self-rated health is a complex phenomenon, which is considered by many as a scale where the categories imply continuity. Others view self-rated health as intrinsically distinct health states, which are predicted by different factors.

The evidence provided in this thesis suggest that depending on how it is viewed, different ordinal models can be used to summarise the results. Unlike any other regression model, an ordinal analysis has the capacity to determine the type of covariates which predict health status as a continuum or otherwise.

A single odds ratio estimated over the categories, implies that the predictor behaves in a similar fashion over the ordinal response categories. Thus, the constant slope adjacent category model would be appropriate for summarising the results where health status is considered to be on a continuum. The different slopes models (polytomous, adjacent category or the stereotype) can be applied with the view that health status is not a continuum and that there are certain strong predictors for, say, '*excellent*' and '*good*' categories and another set of strong predictors for '*fair*' and '*poor*' categories. The partially constrained adjacent category model can be used where the covariates demonstrate both types of effects (constant and different slopes).

8.1.9 Sparse data

Physical disability and cognitive decline in elderly subjects (particularly those with poor physical and mental health) often means that this sub-population results in poor compliance in research studies (see section 2.3.3). Thus sparse data are more likely to occur in studies where the population is elderly people. The data distribution of the Townsend disability score was generally positively skewed with sparse data occurring in the upper end of the scale (although for the very old respondents, thee distribution was more negatively skewed). One would envisage that this would be a typical distribution given the type of study population we are presented with. The sparse data are problematic (and result in poor fit models) when a category where such data are present is used as a single data point in the formation of the logits and the difference between the marginal probabilities of this category and the other (i.e. used in the numerator and denominator of the logit formation) is relatively large.

8.2 Conclusion

8.2.1 Addressing the hypotheses of this thesis

At the end of Chapter 1 the aims of this thesis together with related hypotheses were stated. Given the results and findings, the following is specified.

Hypothesis 1: Evidence in this thesis suggested that this hypothesis was true.

Ordinal regression models compared to linear and binary logistic models are indeed the most appropriate methods for analysing ordinal data as measured on quality of life scales in elderly people.

Hypothesis 2: There was enough evidence to refute this hypothesis.

The evidence in this thesis indicated that the stereotype model was of little use in the context of assessed ordinal quality of life scales.

8.2.2 Contribution to the Literature

The strengths of this study are as follows:-

Health of Elderly people

Results indicate that older people vary in their ratings of their health and that the demographic characteristics – occurrence of heart attack and smoking predict these variations. Self-rated good health was associated with non-smokers and those who had not had a heart attack. The ordinal regression analysis provided further insight into assessing the health of elderly people. This analysis showed that having a heart attack affected the health of an elderly person more than if he/she was a smoker (see Table 8.1). The effect of 'heart attack' is more pronounced for those elderly people with '*good*' compared to '*excellent*' or '*fair*' compared to '*good*' health than for those with '*poor*' compared to '*fair*' health.

Regardless of sex and age-group, older people who have had more than 13 years full-time education are more likely to have a better form of disability than those with less than 13 years full-time education. There is implication that although 'full-time education' is a predictor of physical disability, it is not a very strong predictor.

Statistical Contribution

The statistical contributions were :-

- Highlighting the inadequacy of linear regression models for analysing ordinal quality of life data collected in elderly subjects;

- Illustrating the limitation of binary logistic models for analysing ordinal quality of life data;
- Offering statistical methodology (based on first principles) that can be used to fit all ordinal regression models. These methods can be easily implemented in *SAS*. It is possible that these methods can be implemented in other software packages, if the facilities to compute the design matrix and the different type of logits are available. These statistical methods allow one to fit any ordinal regression model and constrain the parameters in any way possible.

  In the literature, the partial proportional odds models could be fitted with either all the covariates having proportional odds or all the covariates having partial proportional odds (Peterson and Harrell, 1990). The facility to fit covariates with both proportional odds and partial proportional odds has not been available previously. Also, the programs to fit the unconstrained/partially constrained continuation ratio models require manipulation of the data (Cole and Ananth, 2001). The methods in this thesis do not require the user to change the format of the data and therefore are easily implemented to fit these models. In addition, these methods have allowed one to assess the goodness-of-fit and residuals of the non-linear version of the stereotype model.

- Illustrating the limitations of the constrained partial proportional odds model.

- Identifying the particular features of different ordinal regression models (constant, different slopes models and partially constrained models);

- Other modelling aspects such as residual analysis, sparse data and interpretation of interaction terms that are particularly relevant when presented with quality of life data on elderly subjects have been assessed.

Medical Contribution

- Unlike any previous study given in the literature, this study illustrates that given a discrete assessed ordinal quality of life scale, (such as the Health Status index), the most appropriate model for analysing the data is the partially constrained Adjacent Category Model. This result can be generalised to other assessed ordinal scales.

- Given a group continuous scale, the results from the analysis suggest that the fully constrained Continuation Ratio model would be the 'best' model to fit the data. However, one has to be cautious in generalising these results to other multi-item quality of life scales, as the assumption of constant slope may not be satisfied by the covariate of interest (although it may be 'forced' upon the adjusting covariates). In the light of this, and the other results in this thesis, it is recommended that other possible models which may be more likely to fit the data and therefore are adequate include (i) the unconstrained Partial Proportional Odds model; (ii) the partially constrained Continuation Ratio models.

- Sparse data are more likely to occur given that our study population are elderly people (see section 2.3.3). Also, the data distribution of the Townsend disability score is typical of what one might expect given our study population. Ordinal regression models are sensitive in detecting categories where sparse data are present and these data are highlighted as outliers in the residual analysis.

Health Status

Health status has been identified as a complex phenomenon in the literature (see Chapter 3). Certain investigators view health status as implying continuity and the border separating 'good' from 'bad' health is vague. Others have shown that there are different predictors for good and poor health. As the research community stands divided over how to define health status, there is a need to accommodate the different ways that health status is perceived. Ordinal regression models offer the investigator (i) assessing the effect of the predictors over the entire health status scale, assuming it implies continuity (using constant slope adjacent category model), or alternatively (ii) ways of identifying strong predictors for the different categories of health status (e.g. fitting different slope adjacent category model). The partially constrained models offer the option of modelling covariates that may be a mixture of those which assume health status as continuous or otherwise.

Townsend Disability Score data

In Chapter 7, it was found that if the constant slope assumption was satisfied (as it was for 'full-time education'), then the binary logistic regression model as fitted by McGee et al. (MRC CFAS[1] study (1998)) was an appropriate method for analysing the data. In this thesis, the fully constrained continuation ratio model was recommended as the 'best' models for describing the data. This model provided identical odds ratios to the latter binary model. This suggested that the binary model as fitted by McGee et al. (MRC CFAS[1] study (1998)) was a satisfactory method. However, in the latter publication, nine covariates were used in the model with each covariate providing an adjusted odds ratio. There is a possibility that for some of the covariates (where one cannot assume a constant effect over the continuous disability scale), the adjusted estimate of the odds ratio may be misleading.

8.2.3 Limitations of the Study

The methodological limitations of this study are now provided along with some indication of the size of these limitations within the framework of ordinal regression models.

- *Age group covariate in the Townsend Disability Score data as ordered categorical covariate:* The age group covariate was fitted as an ordered categorical covariate in the analysis (with the exception of the binary model). This meant that the age group categories took on a linear structure, and this severely constrained the fit of all the models fitted in the analysis. The reasons age group was taken as ordinal were: (i) that this was its true form and (ii) computationally it was easier to fit and manage. As a result, the full potential of the models (in particular the assessment of the model assumptions and the residual analysis) could not be exploited due to the confounding effect of the ordinal structure imposed on the age group categories. This considerably limited the interpretation of the results. The analysis of this dataset was also carried out using age group as a categorical variable and the results of the 'full-time education' covariate were not very different to those presented in this thesis.

- *Asymptotic Relative Precision:* A general conclusion regarding the Asymptotic Relative Precision (ARP) values was drawn based on the evidence provided. However, the APRs could not be obtained for all the ordinal regression models and therefore these measures were computed only where it was possible. Despite this, the results could be extrapolated to comfortably accommodate all the ordinal regression models.

- *Goodness-of-fit of the models:* In theory, the method devised by Lipsitz (1996) for examining the goodness-of-fit and assessing the residuals is sound. However, in practice its application is difficult, as the design matrix becomes very difficult to implement. This did not hinder the residual analysis as the goodness-of-fit of each ordinal regression model was then based on a statistic derived from the response functions.

- *Residual Analysis:* The residual analysis for some of the ordinal models was carried out using binary regression diagnostics. However, there were other ordinal regression models (e.g. unconstrained partial proportional odds model, constant slope adjacent category model, fully constrained continuation ratio model and the stereotype model) where this was not possible and the residual analysis was based on the visual assessment of the residuals obtained from the logit values. This led to indefinite conclusions being drawn regarding the outliers, as the choice of these observations was not based on any sound statistical evidence. This limited the conclusions of the residual analysis.

- *Akaike Information Criteria (AIC)*: The AIC statistic provides a means of comparing all the ordinal regression models. However, this measure was only obtained for a few of the models where the –2log-likelihood could be derived. For the other models since weighted least squares was the estimation method used, the –2log-likelihood was not available. Where the AIC statistics were present the models were compared. However, not definite conclusion could be drawn regarding comparison of the fit of all ordinal regression models.

- *Covariate adjustment in the Townsend Disability Scale data:* In the analysis carried out by McGee et al. (MRC CFAS[1], 1998) there were nine covariates adjusted for in the analysis. All these covariates could not be adjusted for when fitting the ordinal regression models. This is because the computation of these models becomes very complex (even with only three covariates), particularly when fitting the design matrix. This substantially limited the conclusions drawn, as the results from the publication could not be fully compared those provided in this thesis.

8.2.4 Areas of Further Research

The results and findings of this study raise further questions regarding the statistical aspects of ordinal regression models. Recommendations for future work are now given.

- *Several covariates in an ordinal regression model:* A researcher often requires several covariates in a model (with the possibility that all are strong predictors of the response). Presently an ordinal regression analysis is restricted to three and possibly no more than four covariates in a model, as the design matrix becomes very complex. Further research is required, in terms of computation, so as that it can be possible to fit several covariates in an ordinal regression model.

- *Estimating the constraints for the constrained partial proportional odds model:* This has already been identified as an area of further research (see section 8.3). This area could possibly be taken further by looking at the continuation ratio model/adjacent category model where one could estimate constraints (similar to the ordering constraints used in the non-linear stereotype model). If such constraints simplified the different slopes models, without sacrificing the goodness-of-fit then these would be ideal to use for interpretation of the data.

- *Residual Analysis:* The limitations of the methods used to assess the residuals of some ordinal regression models are discussed above, and there is evidence that this need further research.

- *Indistinguishability in the Stereotype model:* More research is required to compare models where indistinguishability has been imposed and where there are some or no indistinguishable categories. Presently the software to do this is under-developed.

- *Computation of the models*: Particular attention is drawn here to the bootstrapping technique and the computation of the Lipsitz's (1996) statistic. The bootstrapping technique was very computationally intensive and time consuming. Also, the Lipsitz's (1996) statistics could not be computed due to the complex design matrix. These have left gaps for further research.

Engel (1988) mentions working towards 'one super model' that would allow one to compute any type of ordinal regression within its framework. This thesis has certainly provided some steps towards achieving this, with the implementation of the method based on the use of the design matrix. Also, this study has provided a deeper understanding of the analysis of ordinal data (as collected on elderly subjects) using different regression models. Research on health-related quality of life as measured on elderly people is at an early stage of development. Further insight into quality of life can be provided by the use of appropriate statistical methods, such as ordinal regression models. Such insights can guide attempts to intervene most efficiently in encouraging elderly people to alter habits/lifestyles that may be deleterious to their health and quality of life. It is hoped that as the results and findings are disseminated it will encourage investigators to use these models for analysing their data as opposed to other statistical methods (such as the linear or binary regression models as is often done in the literature).

REFERENCE LIST

Agresti A. *Analysis of Ordinal Categorical Data*, Wiley: New York, 1984.

Agresti A. Tutorial on modelling ordered categorical response data. *Psychological Bulletin* 1989; **105**:290-301.

Agresti A. Modelling ordered categorical data: recent advances and future challenges. *Statistics in Medicine* 1999; **18**:2191-2207.

Ananth C, Kleinbaum D. Regression models for ordinal responses: a review of methods and applications. *International Journal of Epidemiology* 1997; **26**:1323-1333.

Anderson J. Regression and ordered categorical variables (with discussion). *Journal of the Royal Statistical Society B* 1984; **46**:1-30.

Armstrong B, Sloan M. Ordinal regression models for epidemiologic data. *American Journal of Epidemiology* 1989: **129**:191-204.

Arnold S. Measurement of quality of life in the frail elderly. In: Birren J, Lubben J, Cichowlas Rowe J, Deutchman D, eds. *The Concept and Measurement of Quality of life in the Frail Elderly*. Academic Press, San Diego: 1991; 50-73.

Ashby D, Pocock S. and Shaper A. Ordered polytomous regression: An example relating serum biochemistry and haematology to alcohol consumption. *Applied Statistics* 1986; **35**:89-301.

Ashford J. An Approach to the analysis of data for semi-quantal responses in biological assay. *Biometrics* 1959; **15**:573-581.

Barsky A, Cleary P, Klerman G. Determinants of perceived health status of medical outcomes. *Social Science and Medicine* 1992; **34**: 1147-1154.

Begg C, Gray R. Calculation of the polytomous logistic regression parameters using individualised regressions. *Biometrika* 1984; **71**:11-18.

Bender R, Grouven U. Ordinal logistic regression in medical research. *Journal of Royal College of Physicians of London* 1997; **31**:546-551.

Bender R, Grouven U. Using binary logistic regression models for ordinal data with non-proportional odds. *Journal of Clinical Epidemiology* 1998; **51**:809-816.

Bender R, Benner A. Calculating ordinal regression models in *SAS* and *SPSS*. *Biometrical Journal* 2000; **42**:677-699.

Bennett S. Analysis of survival data by the proportional odds model. *Statistics in Medicine* 1983; **2**:279-285.

Box G, Tidwell P. Transformation of the independent variables. *Technometrics*1962; **4**:531-550.

Branch L, Jette A, Evasshwick C. Towards understanding elders' health service utilization. *Journal of Community Health* 1981; 7:80-92.

Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* 1990; **46**:1171-1178.

Cassel C. Researching the health needs of elderly people. *British Medical Journal* 1994; **308**:1655-1656.

Cohen M, Tell E, Wallack S. Client-related risk factors for nursing home entry among elderly adults. *Journal of Gerontology* 1986; **4**:785-792.

Cohn J, Sugar J. Determinants of quality of life in institutions: perceptions of frail older residents, staff and families. In: Birren J, Lubben J, Cichowlas Rowe J, Deutchman D, eds. *The Concept and Measurement of Quality of life in the Frail Elderly.* Academic Press, San Diego: 1991; 28-49.

Cole S. A reanalysis of data from 'Regression models for ordinal response'. *International Journal of Epidemiology* 1999; **28**: 805.

Cole S, Ananth C. Regression models for unconstrained, partially or fully constrained continuation odds ratios. *International Epidemiological Association* 2001; **30**:1379-1382.

Cox C, Chuang C. A comparison of chi-square partitioning and two logit analyses of ordinal pain data from a pharmaceutical study. *Statistics in Medicine* 1984; **3**: 273-285.

DeMaris A. A framework for the interpretation of first-order interaction in logit modelling. *Psychological Bulletin* 1991; **110**:557-570.

Department of Health. *NHS R&D Strategic Review. Aging and Age-associate disease and disability.* Report of Topic Working Group June 1999.

Department of Health. *National Service Framework for Older People: modern standards and service models.* Department of Health March 2001: 136-150.

Ebrahim S, Wallis C, Brittis S, Harwood R, Graham N. Long-term care for elderly people. *Quality Health Care* 1993; **2**: 198-203.

Engel J. Polytomous logistic regression. *Statistica Neerlandica* 1988: **42**:233-52.

Farquhar M. Elderly people's definition of quality of life. *Social Science and Medicine* 1995; **41**:1439-1446.

Feinberg B. *Analysis of Cross-classified Data.* Second Edition. Cambridge: Massachusetts Institute of Technology Press, 1980.

Fitzpatrick R, Davey C. Assessing quality of life. In: Evans J, Williams T, Beattie B, Michel J and Wilcock G. eds. *Oxford Textbook of Geriatric Medicine.* Oxford University Press: 2000; 1147-1153.

Fletcher A, Dickinson J, Philp I. Review: audit measures: quality of life instruments for everyday use with elderly patients. *Age and Ageing* 1992; **21**:142-150.

Folstein M, Folstein S, McHugh P. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975; **313**:1419-1420.

Fylkesnes K, Forde O. The Tromso study: Predictors of self-evaluated health – has society adopted the expanded health concept? *Social Science and Medicine* 1991; **32**:141-146.

Fylkesnes K, Forde O. Determinants and dimensions involved in self-rated evaluation of health. *Social Science and Medicine* 1992; **35**: 271-279.

GAUSS Programming System, version 3.0: QUANTAL Applications Module. Aptech Systems, Kent, WA, 1992.

Gill P, Murray W. Quasi-Newton methods for unconstrained optimisation. *Journal of the Institute of Mathematics and its Applications* 1972; **9**:91-108.

Goodman L. Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of American Statistical Association* 1979, **74**:537-552.

Goodman L. The 1983 Henry L. Rietz memorial lecture. The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models and asymmetry models for contingency tables with or without missing entries. *Annual Statistics* 1985; **13**:10-69.

Greenland S. Alternative models for ordinal logistic regression. *Statistics in Medicine* 1994; **13**:1665-1677.

Greenwood C, Farewell V. A comparison of regression models for ordinal data in an analysis of transplanted-kidney function. *The Canadian Journal of Statistics* 1988; **16**:325-335.

Grundy E. Demography and old age. *Journal of American Gerontological Society* 1983; **31**: 325-332.

Grundy E. The Epidemiology of Aging. In: Schroots J, eds. *Textbook of Geriatrics and Gerontology.* 4th Edition. Churchill and Livingstone Publication. Edinburgh, 1992.

Grundy E. Population Review: (5) The population aged 60 and over. *Population Trends* 1996; **84**:14-20.

Grundy E. The health and health care of older adults in England and Wales 1841-1994. In: Charlton J, Murphy M, eds. *The Health of Adult Britain 1984-1994.* London: Stationery Office, 1997: 182-203.

Grundy E. The Epidemiology of Aging. In: Tallis R, Fillit H, Brocklehurst J, eds. *Brocklehurst's Textbook of Geriatric Medicine and Gerontology.* 5th Edition. Churchill Livingstone, 1998.

Guisan A, Harrell F. Ordinal response regression models in ecology. *Journal of Vegetation Science* 2000; **11**:617-626.

Harrell F, Margolis P, Gove S, Mason K, Mulholland E, et al. Tutorial in Biostatistics: Development of a clinical prediction model for an ordinal outcome: The World Health Organisation multi-centre study of clinical signs and etiological agents of pneumonia, sepsis and meningitis in young infants. *Statistics in Medicine* 1998; **17**: 909-944.

Hastie T, Botha J, Schnitzler C. Regression with an ordered categorical response. *Statistics in Medicine* 1989; **8**:785-794.

Hays J, Schoenfeld D, Blazer D. Determinants of poor self-rated health in late life. *American Association for Geriatric Psychiatry* 1996; **4**:188-196.

Hendrickx J. Special restrictions in multinomial logistic regression. *Stata Technical Bulletin* 2000; **STB-56**:18-26.

Holtbrügge W, Schumacher M. A comparison of regression model for the analysis of ordered categorical data. *Applied Statistics* 1991; **40**:249-259.

Hosmer D, Lemeshow S. A goodness-of-fit test for multiple logistic regression model. *Communications of Statistical Theory and Methods* 1980; **10**:1043-1069.

Idler E, Angel R. Self-rated health and mortality in the NHANES-1 epidemiologic follow-up study. *American Journal of Public Health* 1990; **80**; 446-452.

Idler E, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behaviour* 1997; **38**:21-37.

Institute of Medicine. *Health outcomes for older people.* 1996 National Academy Press, Washington, DC.

Jagger C, Clarke M. Mortality risks in the elderly: five-year follow-up of a total population. *International Journal of Epidemiology* 1988; **17**: 111-134.

Joffe M, Greenland S. Standardized estimates from categorical regression models. *Statistics in Medicine* 1995; **14**:2131-2141.

Jorm AF, Korten AE, Henderson AS. The prevalence of dementia: a quantitative integration of the literature. *Acta Psychiatrica Scandinavica* 1987; **76**:465-479.

Kaplan G, Barell V, Lutzky A. Subjective state of health and survival in elderly adults. *Journal of Gerontology and Social Science* 1988; **43**:114-120.

Khaw K-T. How many, how old, how soon? *British Medical Journal* 1999; **319**: 1350-1352.

Kivinen P, Halonen P, Eronen M, Nissinen A. Self-rated health, physician-rated health and associated factors among elderly men: the Finnish cohorts of the seven countries study. *Age and Ageing* 1998; **27**: 41-47.

Koch G, Amara I, Singer J. A two-stage procedure for the analysis of ordinal categorical data. In: P.K. Sen, eds. *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences*. North-Holland, Amsterdam: 1985; 357-387.

Läärä E, Matthews J. The equivalence of two models for epidemiologic data. *Biometrika* 1985; **72**:206-207.

Lee J. Cumulative logit modelling for ordinal response variables: application to biomedical research. *Computer Applications in Biosciences* 1992; **8**:555-562.

Lesaffre E. Logistic discriminant analysis with applications in electrocardiography. Unpublished D. Sc. Thesis 1986, University of Leuven, Belgium.

Lipsitz S. Goodness-of-fit tests for ordinal response regression models. *Applied Statistics* 1996; **45**:175-190.

Lu M. Determinants of residential satisfaction: ordered logit vs. regression models. *Growth and Change* 1999; **30**:264-287.

Lundh U, Nolan M. Ageing and quality of life 1: towards a better understanding. *British Medical Journal of Nursing*, 1996; **20**: 1248-1251.

Lunt M. Stereotype Ordinal Regression. *Stata Technical Bulletin* 2001; **STB-61**:12-18.

Mackenbach J, Van Den Bos J, Joung I, Mheen H, Stronks K. The determinants of excellent health: Different from the determinants of ill-health? *International Journal of Epidemiology* 1994; **23**:1273-1281.

Mackenbach J, Kunst A, Careleers A, Groenhof F, Guerts J. Socioecomonic inequalities in morbidity and mortality in Western Europe. *Lancet* 1997; **349**: 1655-1659.

Manderbacka K, Lahelma E, Maritainen P. Examining the continuity of self-rated health. *International Epidemiological Association* 1998; **27**: 208-213.

Manor O, Matthews S, Power C. Dichotomous or categorical response? Analysing self-rated health and lifetime social class. *International Journal of Epidemiology* 2000; **29**:149-157.

Mantel N. Chi-squared tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association* 1963; **58**:690-700.

McCullagh P. Regression models for ordinal data (with discussion). *Journal of the Royal Society of Statistics B* 1980; **42**:109-42

McCullagh P, Nelder J. *Generalised Linear Models*. 2nd Edition, Chapman and Hall, New York, 1989.

McKelvey C, Zavoina W. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 1975; **4**:103-120.

Mossey J, Shapiro E. Self-rated health: A predictor of mortality among the elderly. *American Journal of Public Health* 1982; **72**:800-808.

MRC CFAS. Socioeconomic status and the expectation disability in old age: estimates for England. *Journal of Epidemiology and Community Health* 2000; **54**:286-292.

MRC CFAS[1]. The description of activities of daily living in five centres in England and Wales. *Age and Ageing* 1998; **27**:605-613.

MRC CFAS[2]. Cognitive function and dementia in six areas of England and Wales: the distinction of MMSE and prevalence of GMS organicity levels in the MRC CFAS study. *Pyschological Medicine* 1998; **28**:319-335.

O'Mahoney P, Rodgers H, Thomas R, Dobson R, James O. Is the SF-36 suitable for assessing health status of older stroke patients ? (form used for self-reporting on health status in the United Kingdom). *Age and Ageing* 1998: **27**: 19-22.

Parmar M, Machin D. *Survival Analysis: A practical approach.* Chichester, Wiley 1995.

Peterson B, Harrell F. Partial proportional odds model and the LOGIST procedure. Paper presented at *SAS* Users Group International, thirteenth annual conference. Orlando, FL, 1988.

Peterson B, Harrell F. Partial proportional odds models for ordinal response variables. *Applied Statistics* 1990: **39**:205-17.

Peterson B, Harrell F. Proportional Odds Model (Correspondence). *Biometrics* 1992; **48**: 325-326.

Pettinger N. NHS costs. Age-old Myths. *Health Service Journal* 1998; **108**: 24-25.

Pijls L, Feskens E, Kromhout D. Self-rated health, mortality and chronic diseases in elderly men: The Zutphen study, 1985-1990. *American Journal of Epidemiology* 1993; **138**: 840-848.

Powell Lawton M. A Multidimensional view of quality of life in frail elders. Quality of life index. In: Birren J, Lubben J, Cichowlas Rowe J, Deutchman D, eds. *The Concept and Measurement of Quality of life in the Frail Elderly.* Academic Press, San Diego: 1991.

Power C, Matthews S, Manor O. Inequalities in self-rated health: explanations from different stages in life. *Lancet* 1998; **351**:1009-1014.

Rao C. Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proceedings Cambridge Philological Society* 1947; **44**:50-57.

Rao C. *Linear Statistical Inference and its Applications.* 2nd Edition, New York: Wiley, 1973.

Robinson L, Jewel N. Some surprising results about covariate adjustment in Logistic Regression models. *International Statistical Review* 1991; **58**:227-240.

SAS supplemental Library user's guide (1986), *SAS* Institute.

SAS/STAT User's Guide, Version 5.18. Cary, NC: *SAS* Institute.

SAS/STAT User's Guide (1986) Cary, NC: *SAS* Institute.

SAS/STAT User's Guide (1997) Cary, NC: *SAS* Institute.

SAS/STAT User's Guide, Version 6, 4th Edition, Cary, NC: *SAS* Institute.

Scott S, Goldberg M, Mayo N. Statistical assessment of ordinal outcomes in comparative studies. *Journal of Clinical Epidemiology* 1997; **50**:45-55.

Seigel S, Castellan N. *Non-parametric Statistics for the Behavioural Sciences*. Second Edition. McGraw-Hill International Editions, 1988.

Smith A, Shelley J, Dennerstein L. Self-rated health: biological continuum or social discontinuity. *Social Science and Medicine* 1994; **39**:77-83.

Snell E. A scaling procedure for ordered categorical data. *Biometrics* 1964; **20**:592-607.

STATA 3.0 Computing Resource Centre, Santa Monica, 1992.

STATA manual, Version 7. College Station, TX: Statacorp, 2001.

Stokes M, Davis C, Koch G. Categorical data analysis using the *SAS* system, Cary, NC: *SAS* Institute, 1995.

Strömberg U. Collapsing ordered outcome categories: A note of concern. *American Journal of Epidemiology* 1996; **144**:421-424.

Stucki G, Daltroy L. Interpretation of change scores in ordinal clinical scales and health status outcomes: the whole may equal the sum of the parts. *Journal of Clinical Epidemiology* 1996; **49**:711-717.

Swerdlow, dos Santos Siva. *Atlas of Cancer Institute in England and Wales 1968-1985*. Oxford University Press, Oxford, 1993.

Tallis R. Rehabilitation of the elderly in the 21st century. *Journal of the Royal College of Physicians of London* 1992; **26**:413-422.

Townsend P. *Poverty in the United Kingdom*. Harmonds-worth, Pelican, 1979.

Van Duiju L, Hofman C. Risk factors for Alzheimer's disease: the EURODERM collaborative re-analysis of case-control studies. *Neuroepidemology* 1992; **11**:106-113.

Walker S, Duncan D. Estimation of the probability of an event as a function of several independent variables. *Biometrika* 1967: **54**: 167-79.

Walters S, Campbell M, Lall R. Design and analysis of trials with quality of life as an outcome: A practical guide. *Journal of Biopharmaceutical Statistics* 2001; **11**:155-176.

Ware J, Gandek B. Overview of the SF-36 health survey and the International Quality of life assessment (IQOLA) project. *Journal of Clinical Epidemiology* 1998; **51**:903-912.

Wilkin D, Hughes B. The elderly and the health service. In: Phillipson C, Walker A, eds. *Ageing and Social Policy: A critical assessment*. Gower, Aldershot, 1986.

Wolinsky F, Arnold C. A different perspective on health and health services utilisation. *Annual Review of Gerontology and Geriatrics* 1988; **8**:71-101.

Worsley A. Layperson's evaluation of health: an exploratory study of an Australian population. *Journal of Epidemiology and Community Health* 1990; **44**: 7-11.

Yesavage J, Brink T, Rose R, Lum O. et al. Development and evaluation of the Geriatric Depression Screening Scale: A preliminary report. *Journal of Psychiatric Research* 1983; **17**:37-49.

Zigmond A, Snaith R. The Hospital Anxiety and Depression Scale. *Acta Pyschiatria Scandinavica* 1983; **67**: 361-370.

APPENDIX I - ORDINAL SCALES AND COVARIATES

Listing Ia: Ordinal scale variables from the Core MRC CFAS Prevalence Screening Questionnaire.

V33. How often see relative to talk to:
0. Never 1. Daily 2. 2-3 times a week 3. At least weekly 4. at least monthly 5. less often 7. don't know 8. not asked

V35. How often do you see the relations which you have most contact with :
1. daily 2. 2-3 times a week 3. at least weekly 4. at least monthly 5. less often 7. don't know 8. not asked

V37. How often do you see your neighbours?
0. Never/no neighbours 1. daily 2. 2-3 times a week 3. at least weekly 4. at least monthly 5. less often 7. don't know 8. not asked

V40. Subject rating of own health?
0. Excellent 1. Good 2. Fair 3. Poor 7. don't know 9. not asked

V108. Memory difficulty – is this a problem for you?
0. No 1. Yes, moderate 2. Yes, severe

V110. Memory difficulty – have you tended to forget things recently?
0. No 1. Yes, several time a week 3. Yes at least daily

V111. Memory difficulty – what kind of things do you forget? Names of family and close friends
0. No 1. Yes, several times a week 2. Yes, at least daily

V112. Memory difficulty – what about where you put things?
0. No 1. Yes, several times a week 2. Yes, at least daily

V 114. Memory loss – when did you first notice this beginning?
1. less than 1 year ago 2. In the last 1-2 years 3. In the last 3-4 years 4. In the last 5-10 years 5. over ten years ago 8. no answer 9. not asked

V117. Ever had problems sleeping?
0. Never 1. Seldom 2. Sometimes 3. Often 4. All the time 8. No answer 9. Not asked

V120. Do you snore?
1. No 2. Sometimes 3. All the time

Listing Ib: Disability Scale (from which the ordinal Townsend Disability Scale is derived)

Q 121. I would like to know if you are able, or if you have any difficulty with, the following ten activities. So are you able to cut your own toe-nails?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Q122. Are you able to wash all over or bath?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Q123. Are you able to get on the bus?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Q124. Are you able to go up and down stairs?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Q125. Are you able to do the heavy housework?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Q126. Are you able to shop and carry heavy bags?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Q127. Are you able to prepare and cook a hot meal?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Q128. Are you able to reach an overhead shelf?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Q129. Are you able to tie a good knot in a piece of string?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Q130. Are you able to put on your shoes and socks or stockings?

0. (No) need help	1. (Yes), some difficulty	2. (Yes), no difficulty
7. Don't know	8. No answer	9. Not asked

Table 1a: The covariates and their description chosen for the Health Status data

Covariate	Description
Sex	males/females and there were no missing data on this variable.
Centre	Cambridge, Gwynedd, Newcastle, Nottingham and Oxford and no missing data were recorded on this variable.
Age-group	The continuous form of the age covariate was taken and age groupings as specified in the MRC CFAS[1] (1998) study formed. These groups were <70, 70-74, 75-79, 80-84, ≥ 85 years and no missing data were recorded on this variable.
Full-time education	The level of education was recorded as a continuous variable on the database, and as a result, this variable was categorised using groups presented in the MRC CFAS[1] (1998) study. The categories were less than 9 years, equal to 9, 10 to 12 and 13 or more years full-time education and there were no missing data on this variable.
Marital Status	Marital status was recorded using five categories: married; cohabiting; single; widowed; divorced. In addition to these, respondents who did not answer, or where the question was not asked, were also recorded and on the database a total of 4 (0.0003%) respondents did not answer this question. Thus there were 12618 (99.9%) subjects with available data on health status and marital status.
Smoking	The question 'Do you smoke?' was asked and a response from 'No', Yes', 'missing', 'no answer' and 'not asked' was recorded. There were 78 (0.006%) respondents who had missing data with regards to smoking/not smoking and 1 subject was not asked the question. In total 12543 subjects has an assessment of health status and smoking.
Alcohol Intake	The question that seemed most appropriate from the database relating to alcohol was 'Have you ever taken an alcoholic drink of any kind?' to which a response of 'Yes', 'No', 'not asked' and 'no answer' was given. There were 88 (0.007%) respondents who either had missing data, or did not answer or were not asked the question. The remaining 12534 (99.3%) respondents had had alcoholic drink of some kind.
Occurrence of Heart attacks	The question 'Have you ever suffered from a heart attack?' was asked for assessing the occurrence of heart attacks,. The response was recorded as 'Yes', 'No', 'No answer', 'Not asked' and 'missing'. It was evident that 24 (0.002%) subjects had missing data with regards to occurrence of heart attack, 5 subjects did not answer and 1 subject was not asked the question. In total 12595 (99.8%) respondents had a yes/no response.
Angina	The question 'Have you ever suffered from angina?' was asked and a response was recorded from 'No', 'Yes', 'No answer' and 'Not asked'. There were 7 missing assessments on this variable and 7 more respondents did not answer the question and 2 respondents were not asked. In total 12606 (99.9%) of the subjects had valid assessments regarding angina.
Sugar Diabetes	An assessment of diabetes was also made and recorded on the database as 'Have you ever had sugar diabetes?' and a response of 'No', 'Yes', 'No answer' and 'Not asked' was made. There were 32 (0.003%) subjects who either did not answer or were not asked the question or there were missing data. From the remaining 12590 (99.7%) subjects a 'Yes/No' response was recorded regarding sugar diabetes.

Table Ib: The p-values from the Pearson's chi-squared tests of association

	Health Status	*Centre*	*Sex*	*Age group*	*Education*	*Marital Status*	*Smoking*	*Alcohol*	*Heart Attack*	*Angina*	*Sugar diabetes*
Centre	<0.0001	-	-	-	-	-	-	-	-	-	-
Sex	<0.0001	0.0004	-	-	-	-	-	-	-	-	-
Age	<0.0001	<0.0001	<0.0001	-	-	-	-	-	-	-	-
Education	<0.0001	<0.0001	0.0017	<0.0001	-	-	-	-	-	-	-
Marital Status	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-	-	-	-
Smoking	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-	-	-
Alcohol	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-	-
Heart Attack	<0.0001	0.0101	<0.0001	0.3486	0.5477	<0.0001	0.9821	0.0053	-	-	
Angina	<0.0001	<0.0001	<0.0001	0.0002	0.2872	0.0028	0.0002	0.1921	<0.0001	-	-
Sugar Diabetes	<0.0001	<0.0001	0.0005	0.0016	0.0692	0.220	0.0713	0.0054	<0.0001	<0.0001	-

APPENDIX II - COMPUTATION OF THE REGRESSION MODELS

Section 1: Computation of the regression models

Linear regression models: The linear regression models were fitted using PROC GLM in *SAS*, as the response was taken to be on a continuous scale.

Binary logistic regression models: PROC LOGISTIC can be used to fit binary logistic regression model as described by McGee et al. (MRC CFAS[1], 1998). This procedure is unable to allow for the ordinal nature of 'age group' and therefore this covariate had to be fitted as categorical. All other binary logistic regression models were fitted using PROC CATMOD to keep in consistent with the ordinal regression models.

Polytomous Models: The procedure used to fit the polytomous model was PROC CATMOD and maximum likelihood estimates were provided.

Proportional odds models: PROC LOGISTIC provides maximum likelihood estimates for the proportional odds model. It also provides the global χ^2-score statistic that tests all the covariates for the proportional odds assumption. For this thesis, the proportional odds model was fitted in PROC CATMOD, where the estimates were obtained using weighted least squares method. This was the preferred procedure as it was consistent with the other models.

Partial Proportional odds models: The only procedure that could fit the partial proportional odds models was PROC CATMOD. This procedure provided weighted least square estimates.

Adjacent Category Models: The adjacent category models were fitted using PROC CATMOD. This procedure provided weighted least square estimates. The different slopes model was used to test the assumption of a constant slope and some slopes being constrained as equal by the specification of the CONTRAST statement in PROC CATMOD.

Continuation Ratio Models: The constant slope continuation ratio model was fitted in PROC LOGISTIC. This procedure provided a global test of parallel slopes for all the covariates in model. PROC LOGISTIC could have been used to fit the separate binary logistic models based on the cut-points of the continuation ratio model. PROC CATMOD was used instead, to keep in consistent with the other models. Also the unconstrained and partially constrained continuation ratio models were fitted using PROC CATMOD with a specification of the response function and design matrix. This method provided weighted least square parameter estimates. The unconstrained continuation ratio model provides a means for testing for constant slopes.

Stereotype Model: The stereotype model (non-linear) where the constraints were estimated as parameters in the model was fitted using especially devised macros in *SAS* and *STATA* by Hendrickx (2000) and Lunt (2001). The stereotype model (linear form) with the estimated constraints used as constants was fitted using PROC CATMOD.

Section 2: Model fitting

I. Observed sample data

The starting point of the model fitting procedure was to compute the observed sample marginal probabilities (p_{ij}) and the observed sample logits or *response functions* $F(p)$.

(a) Observed sample marginal probabilities

From Table 3.1 $p_i' = (p_{i1}, p_{i2}, \ldots\ldots p_{ic})'$ with $\sum_j p_{ij} = 1$, (where $i = 1\ldots\ldots r$, and $j = 1\ldots\ldots c$) denotes the conditional distribution of Y at level i of the sub-populations (obtained from the levels of the explanatory variables) and $p = (p_1', p_2', \ldots\ldots p_c')$. Then p_{ij} corresponds to the proportion of subjects in each group response and is written as:

$p_{ij} = n_{ij} / n_{i+}$ where n_{ij} is the number of subjects in the i^{th} group or cell who have the j^{th} response. Thus

$p_i' = (p_{i1}, p_{i2}, \ldots p_{ic})$ and p_i' is of dimension (1 x c)

$$\text{and } p' = (p_1', p_2', \ldots p_r') \qquad \text{(A1)}$$

is of dimension (1 x rc).

The rows of the contingency table are considered to be simple random samples from the multinomial distribution; since the rows are independent, the entire table is distributed as product multinomial. Then the covariance matrix for the proportions is the sample estimate covariance matrix for V(p) (of dimension rc x rc) i.e.

$$V(p_i) = \frac{1}{n_{i+}} \begin{bmatrix} p_{i1}(1-p_{i1}) & -p_{i1}p_{i2} & \ldots\ldots\ldots\ldots & -p_{i1}p_{ir} \\ -p_{i2}p_{i1} & p_{i1}(1-p_{i2}) & \ldots\ldots\ldots\ldots & -p_{i2}p_{ir} \\ -p_{i3}p_{i1} & -p_{i3}p_{i2} & p_{i3}(1-p_{i3})\ldots\ldots & -p_{i3}p_{ir} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ -p_{ir}p_{i1} & -p_{ir}p_{i2} & \ldots & p_{ir}(1-p_{ir}) \end{bmatrix} \qquad \text{(A2)}$$

Then the covariance matrix for the entire table can be written as:

$$V(p) = \begin{bmatrix} V_1 & 0 & 0\ldots\ldots & 0 \\ 0 & V_2 & 0\ldots\ldots & 0 \\ 0 & 0 & V_3\ldots\ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & V_r \end{bmatrix} \qquad \text{(A3)}$$

Observed sample response functions and their variance-covariance matrix

Once the proportion vector and covariance matrix have been computed, the observed sample response functions are obtained.

Let $F(p)$ be a vector of $u \leq r(c-1)$ response functions. Then $[F(p)]' = [f_1(p), f_2(p), \ldots f_u(p)]$ and $f_m(p)$ are functions of the elements of p_{ij} that are assumed to have continuous second-order partial derivatives.

Then let $Q(p) = \left(\frac{\partial f_m(p)}{\partial p_{ij}} \right)$ which is a u x cr matrix, with m = 1,.... u and for all combinations (i, j). We assume that the $f_m(p)$ are linearly independent, so that Q has rank u. Logit response functions have the form

$$F(p) = K\log(Ap) \qquad \text{(A4)}$$

for certain matrices K and A, where log transforms a vector to the corresponding vector of natural logarithms. In this case $Q(p) = KD^{-1}A$, where D is a diagonal matrix with the elements of the vector $A\pi$ on the diagonal.

The asymptotic sample variance of $F(p)$ is

$$S = Q(p)V(p)Q(p)'. \qquad \text{(A5)}$$

The above provides us with the observed probabilities and response functions. The idea behind model fitting is to use variation on the design matrix X and provide estimates of the parameters that lead to predicted probabilities and logits. Thus, the logit models can be expressed in the form:

$$F(\pi) = X\beta \qquad \text{(A6)}$$

where $F(\pi)$ is the vector of the $r(c-1)$ logits, X is a $r(c-1)$ x v design matrix and β is a vector of the parameters $\beta_1,.....,\beta_v$ to be estimated.

<u>II. Parameter estimation and Inference</u>

Two methods are used to estimate the parameters in a model:

(a) the maximum likelihood method estimates the parameters of the linear model so as to maximise the value of the joint multinomial likelihood function of the responses.
(b) the weighted least squares method minimises the weighted residual sum of squares for the model. The weights are contained in the inverse of the covariance matrix (A5) of the response functions. Weighted least squares is a generalisation of ordinary least squares that give relatively more weights to a sample logit as its variance decrease.

The following section details these two methods in relation to ordinal regression models.

(a) Maximum likelihood estimation

When presented with logits that are of a generalised form, as in the case of the separate binary logistic, polytomous, continuation ratio (using separate binary logistic regression models) and the linear stereotype models, the analysis is based on maximum likelihood estimation and the Newton-Raphson algorithm is used. These multinomial models can be treated as a direct extension of the binary logistic regression model, and therefore take the response as being nominal. For these models the log-likelihood is concave and parameter estimates necessarily exist and are unique and finite if all observed cell counts are positive. The mechanics of the maximum likelihood estimation method and model fitting for the latter logit regression models is detailed below.

To use the method of maximum likelihood, we start with the observed sample probabilities (p_{ij}) and obtain the log-likelihood.

Generally, the contribution from a single multinomial observation $\{n_{i1},\ldots n_{ic}\}$ to the likelihood function is $p_{i1}^{n_{i1}},\ldots p_{ic}^{n_{ic}}$ and we can write

$$l(p_i; n_i) = \sum_j n_{ij} \log p_{ij} \quad . \tag{A7}$$

The observations and the probabilities are subject to the linear constraints that are non-negative and add to 1.0. Since the N observations are independent by assumption, the total log likelihood is a sum of contributions, one from each of the N observations.

Thus $l(p; n) = \sum_{ij} n_{ij} \log p_{ij}$. (A8)

The first derivative of the log-likelihood can be written as:

$$\frac{\partial \log l(p;n)}{\partial p_{ij}} = \frac{n_{ij} - m_i p_{ij}}{p_{ij}}, \tag{A9}$$

where $m_i = \sum_j n_{ij}$ and is fixed for each i.

The design matrix determines the number and type of parameters that are fitted. The likelihood equations for the parameters are obtained by multiplying (A9), by the derivative of

p_{ij} with respect to each parameter in turn and summing over i and j. For example, suppose we have a generalised ordinal regression model of the form

$$F(\pi_i) = \eta_i = \alpha_j + \sum_{k=1} x_{ik}\beta_{jk} ,$$

then

$$q_r = \frac{\partial \log l(\pi;n)}{\partial \beta_r} = \sum_{ij} \frac{\partial \log l(\pi;n)}{\partial \pi_{ij}} . \frac{\partial \pi_{ij}}{\partial \beta_r}, \qquad r=1,\ldots,v \qquad \text{(A10)}$$

The RHS is made up of two components with the $\frac{\partial \log l(\pi;n)}{\partial \pi_{ij}}$ is obtained from the sample data and $\frac{\partial \pi_{ij}}{\partial \beta_r} = \frac{\partial \pi_i}{\partial \eta_i} . \frac{\partial \eta_i}{\partial \beta_r}$ is obtained partially from the above specified ordinal regression model. The expected value of the second derivative is obtained from (A10) and is used in a series of Newton-Raphson iterations to obtain a sequence of parameter estimate, such that

$$\beta_r^{(t+1)} = \beta_r^{(t)} - (H_r^{(t)})^{-1} q_r^{(t)} \qquad \text{(A11)}$$

where $q_r^{(t)}$ and $H_r^{(t)}$ denote $q_r = 1, 2, \ldots v$ and H_r is the expected values of the second derivative respectively. These are evaluated at estimated generalised probabilities obtained at the t^{th} iteration. Convergence usually occurs within a few cycles and to estimated maximum likelihood parameters and the covariance matrix of the model parameter estimates $V(\beta)$ are produced as a by-product of this method.

The predicted response functions are obtained using the parameter estimates such that $F(\hat{\pi}) = X\beta$ and the estimated covariance-variance matrix is obtained as $\hat{\Sigma} = XV(\beta)X'$.

(b) Weighted least squares method

For the proportional odds, partial proportional odds (constrained and unconstrained), the adjacent category and the continuation ratio models the method of maximum likelihood is not so easily adapted as the response functions are complex functions of the cell counts. Weighted least squares estimation is much easier than maximum likelihood estimation to compute these models.

The weighted least squares estimates of $\beta_1, \ldots\ldots, \beta_v$ is the vector that minimises the quadratic form:

$$Q_w = (F(\pi) - X\beta)'\Sigma^{-1}(F(\pi) - X\beta). \qquad \text{(A12)}$$

This estimate equals

$$\beta = (X'S^{-1}X)^{-1}X'S^{-1}F(p). \qquad \text{(A13)}$$

The covariance matrix for *b* is written as:

$$V(\beta) = (X'S^{-1}X)^{-1}. \qquad \text{(A14)}$$

The design matrix, *X*, is dependent on the type of logits and parameters fitted. The predicted values of the response functions $F(\hat{\pi})$ are smoother than the observed response functions and are obtained by $F(\hat{\pi}) = X(X'S^{-1}X)^{-1}X'S^{-1}F(p)$. The estimated covariance matrix of the predicted response functions is

$$\hat{\Sigma} = X(X'S^{-1}X)^{-1}X'. \qquad \text{(A15)}$$

(c) Statistical Inference

One can address questions about the parameters with the use of the hypothesis tests. Each hypothesis is written in the form:

$$H_0: C\beta=0 \qquad \text{(A16)}$$

and one can investigate whether specified linear combinations of the parameters are equal to zero. The test statistic employed is a Wald statistic that is expressed as:

$$Q_c = (C\beta)'[C(X\hat{\Sigma}^{-1}X)^{-1}C']^{-1}(C\beta) \qquad \text{(A17)}$$

where Q_c is distributed as chi-square with degrees of freedom equal to the number of linearly independent rows in C.

Section 3: Computing the Ordinal Logit Models using SAS

PROC CATMOD in *SAS* was predominantly used to fit the binary and ordinal regression models. Stokes et al. (1995) detail the computation of the polytomous and proportional odds models using methods based on first principles. These methods were adapted to fit the binary, partial proportional odds, adjacent category, continuation ratio (constrained and unconstrained) and the linear stereotype models. The general *SAS* code used in PROC CATMOD was:

```
PROC CATMOD DATA=dataset;
RESPONSE logit/clogit/alogit;
POPULATION X1 X2;
DIRECT X1 X2;
MODEL y-response=design matrix/options;
run;
```

The RESPONSE statement specifies functions of the response probabilities. In the case of the cumulative logits, *CLOGITS* was used; for adjacent category logits, *ALOGITS* was used and if the generalised logits were required *LOGIT* was specified. For the continuation ratio model, no statement exists and one has to compute the response functions according to the *y*-response present in the data (see section 5.2.3.4 ii (d) and section 5.3.3.4 ii (d)).

POPULATION specifies the independent variables that determine the sub-populations that are formed on the basis of cross-classification of the specified covariates. DIRECT allows the specification of the continuous covariates.

The MODEL statement is made up of the dependent variable and the potential sources of variation. These sources of variation were specified in a form of a design matrix.

The code changes depending on the type of response function required and the expression of the systematic component of the model. In the literature cumulative logit models are often modelled using equation (3.8). However, PROC CATMOD specifies the cumulative logits as ln[Pr($Y_i \geq y_j$)/Pr($Y_i < y_j$)]. Also the adjacent category logits are often modelled using equation (3.16) and in *SAS* these logits were based on ln[Pr($Y_i = y_{j-1}$/Pr($Y_i = y_j$)].

There were three scenarios as regards model fitting: (i) a different slopes model fitted where different regression parameters $\{\beta_j\}$ were required (e.g. for the polytomous, different slopes cumulative, different slopes adjacent category and unconstrained continuation ratio models); (ii) a model with constant slopes over all the response categories (e.g. the proportional odds, fully constrained continuation ratio, constant slope adjacent category and the linear stereotype models) , and (iii) a model with a combination of covariates , some of which have different slopes and some of which have constant slopes over the cut-points (e.g. constrained and unconstrained partial proportional odds models and the partially constrained continuation ratio and adjacent category models). For each scenario, the variation was on the design matrix, when fitting the logit models. Details of how the design matrix was constructed are given in Section 5: Statistical Analysis. The appropriate design matrix was taken and the *SAS* code, similar to the one below, was computed.

(a) Scenario 1: Fitting the Polytomous/different slope Cumulative/Unconstrained Continuation Ratio/Adjacent Category models

The general *SAS* code when fitting the different slopes models given the Health Status data was:

```
Proc catmod data =temp2a;
response clogits/logits/alogit;
population  v61_s0_ v150_s0_;
model v40_s0 = (1  0  0  1  0  0  1  0  0,
                0  1  0  0  1  0  0  1  0,
                0  0  1  0  0  1  0  0  1,
                1  0  0 -1  0  0  1  0  0,
                0  1  0  0 -1  0  0  1  0,
                0  0  1  0  0 -1  0  0  1,
                1  0  0  1  0  0 -1  0  0,
                0  1  0  0  1  0  0 -1  0,
                0  0  1  0  0  1  0  0 -1,
                1  0  0 -1  0  0 -1  0  0,
                0  1  0  0 -1  0  0 -1  0,
                0  0  1  0  0 -1  0  0 -1)/pred=prob cov;
contrast 'b11=b21=b31' all_parms 0 0 0 1 -1  0 0  0 0,
                       all_parms 0 0 0 0  1 -1 0  0 0;
contrast 'b21=b22=b32' all_parms 0 0 0 0 0 0 1 -1 0,
                       all_parms 0 0 0 0 0 0 0  1 -1;
contrast 'smoke: 2b11' all_parms 0 0 0  2 0 0  0 0
0/estimate=both;
```

```
contrast 'smoke: 2b21' all_parms 0 0 0  0 2 0  0 0
0/estimate=both;
contrast 'smoke: 2b31' all_parms 0 0 0  0 0 2  0 0
0/estimate=both;
contrast 'h.a.: 2b12' all_parms 0 0 0  0 0 0  2 0
0/estimate=both;
contrast 'h.a.: 2b22' all_parms 0 0 0  0 0 0  0 2
0/estimate=both;
contrast 'h.a.: 2b32' all_parms 0 0 0  0 0 0  0 0
2/estimate=both;
run;
```

Observed marginal probabilities: The marginal probabilities were obtained in *SAS* after the specification of the RESPONSE function and the design matrix, using PROB in the MODEL/OPTIONS statement. In the *SAS* output, the observed probabilities were detailed in the 'Response Probabilities' Table.

Response Functions

In *SAS* the POPULATION statement had the two class covariates 'heart attack' and 'smoke' and no DIRECT statement was specified. Here the 'v61_s0_' is the 'heart attack' covariate and 'v150_s0_' is the 'smoke' covariate. The design matrix was specified in the MODEL statement with the health status ('resp') as the response variable. The RESPONSE statement used one of the logits specified in the code; it depended on which model was being fitted. Given, for example, the polytomous model, where the maximum likelihood estimation method was used the *SAS* output provided the iterations history and the values for the –2 log-likelihood (detailed in the 'Maximum Likelihood Analysis' Table). The maximum likelihood parameter estimates together with their standard errors were obtained in 'Analysis of Maximum Likelihood Estimates' table. The observed and predicted response functions were given in the 'Maximum Likelihood Predicted Values for Response Functions' table. For the different slopes models where the method of estimation was weighted least squares, one required the observed sample response functions and the variance-covariance matrix of the observed sample response functions, *S*. These were displayed in the 'Response Functions and Covariance Matrix' table and the 'Predicted values of the Response Function' table. The standard errors of these response functions were displayed in the 'Predicted values of the Response Functions' table. The predicted response functions and their variance-covariance matrix are also displayed in the same tables.

Contrast Statements : The log odds ratios was obtained by taking the difference in the appropriate response functions as detailed in section 5.2.3.1 (d). The linear combination of the parameters that make up the difference was specified in a

CONTRAST statement and this also provided the estimate of the log odds ratio together with the odds ratios and their 95% confidence intervals. With the specification of /ESTIMATE=BOTH after the CONTRAST statement, a point estimate of $2\beta_{11}$ (i.e. the log odds ratio of smokers versus non-smokers from section 5.2.3.1 (d)) with its standard error was provided.

The CONTRAST statement was also used to test hypotheses such as $H_0: \beta_1 = \beta_2 = \beta_3$. Hypothesis testing was required for assessing the interaction term (where the effects were tested against zero) or the effects used for model assumptions such as the proportional odds or parallel slopes (where the effects were tested for homogeneity). One of the by-products of the contrast statement was the estimate of the odds ratios and their 95% confidence intervals and this provided the relevant summary statistics for many of the models.

(b) Scenario 2: Fitting the Proportional Odd/ constant slope Adjacent Category/constant slope Continuation Ratio/linear Stereotype models

The parameters were constrained to be the same over all the cut-points for the constant slope models. The variation in the above code was in the specification of the design matrix for the proportional odds model. This design matrix was used for the constant slope adjacent category, the constant slope continuation ratio and the linear stereotype models.

Similar *SAS* output to (a) was provided for the response functions and the marginal probabilities.

(c) Scenario 3: Fitting the Partial Proportional Odds models(constrained and unconstrained) models and the partially constrained Continuation Ratio and Adjacent Category models

The design matrix used for the unconstrained partial proportional odds model is detailed in the matrix formulation 5.12 and for the constrained partial proportional odds model is detailed in the matrix formulation 5.17, given the Health Status data. The design matrix for the partially constrained continuation ratio model is detailed in section 5.2.3.4 (iii). Each of these design matrices was incorporated into the above *SAS* code, resulting in weighted least square parameter estimates. The odds ratios and their 95% confidence intervals were obtained using the CONTRAST statement.

Similar *SAS* output to (a) was provided for the response functions and the marginal probabilities.

Section 4: Fitting the Stereotype Models using Hendrickx's (2000) and Lunt's (2001) macros

Macros have been devised in *SAS* and *Stata* by Hendrickx (2000) for fitting the stereotype model using conditional logistic regression. For this thesis, the macros devised in *SAS* were used. The data were transformed into a form suitable for a conditional logistic regression model using the *mclgen* macro. Once the data had been transformed and the model fitted, the same parameter estimates as obtained for multinomial models were estimated. However, in using this method there is an additional advantage in that greater flexibility is allowed such that different constraints can be imposed on the response for different independent variables.

There are two main macros that were used and these are as follows:

(a) *mclgen* –This transformed the data into a person/choice file. In the person/choice file, each respondent had a separate record for each category of the response variable. A stratifying variable (_STRATA) indexed respondents, the response variable indexed response options (_NEWY) and a dichotomous variable (_DEPVAR_) indicated which response option was the respondent's actual choice. The effects of the independent variables were included by creating the y_1 to y_c dummy variables. Suitable transformation of the response dummies allowed for other response functions than the standard logits, in which the highest category was the referent category.

(b) *Mclest* – this entered the dichotomous dependent variable and stratifying variable and then estimated the multinomial model using the conditional logit procedure PROC PHREG. A conditional logit model was characterised by a binary dependent variable, independent variables indicated choices as well as characteristics of the respondents, and a stratifying variable, within which the likelihood was evaluated. In addition, PROC PHREG required a censoring variable (_NOT) in order to estimate conditional logit models. This censoring variable had the mirror value of the new binary dependent variable and was specified in conjunction with this dependent variable. *Mclest* iteratively estimated the multinomial conditional logit model, by first taking the ϕ_j scaling metric as given and estimating the β_k parameters, then taking the β_k parameters and estimating the ϕ_j parameters. This continued until the change in the log likelihood between the successive multinomial logistics models was less than some given value (by default this is equal to 0.0001) or the maximum number of iterations was exceeded (default 10).

This model produced (c-1) standard multinomial intercept parameters, (c-1) independent ϕ_j and a single β for each covariate. In the case where the constraints were estimated, the β_k and ϕ_j parameters were conditional on the estimates and the standard errors of the odds ratios were invalid. Likewise, any inference based on the standard errors was also not correct and instead bootstrap techniques (random sampling with replacement from the original dataset) were applied to obtain the correct standard errors and tests. The aspects of indistinguishability and dimensionality were also assessed using a macro devised in *STATA*, known as *SOREG* (Lunt, 2001).

Table IIa: Log odds predicted for the polytomous model using the Townsend Disability Scale data

Full-time education (f)	**Sex (g)**	**Age group (h)**	$f_{fgh1}(\hat{\pi})$	$f_{fgh2}(\hat{\pi})$	$f_{fgh3}(\hat{\pi})$	$f_{fgh4}(\hat{\pi})$
<13 years	***Male***	***<70***	$\alpha_1+\beta_{11}+\beta_{12}+\beta_{13}$	$\alpha_2+\beta_{21}+\beta_{22}+\beta_{23}$	$\alpha_2+\beta_{31}+\beta_{32}+\beta_{33}$	$\alpha_4+\beta_{41}+\beta_{42}+\beta_{43}$
<13 years	***Male***	***70-75***	$\alpha_1+2\beta_{11}+\beta_{12}+\beta_{13}$	$\alpha_2+2\beta_{21}+\beta_{22}+\beta_{23}$	$\alpha_2+2\beta_{31}+\beta_{32}+\beta_{33}$	$\alpha_4+2\beta_{41}+\beta_{42}+\beta_{43}$
<13 years	***Male***	***75-80***	$\alpha_1+3\beta_{11}+\beta_{12}+\beta_{13}$	$\alpha_2+3\beta_{21}+\beta_{22}+\beta_{23}$	$\alpha_2+3\beta_{31}+\beta_{32}+\beta_{33}$	$\alpha_4+3\beta_{41}+\beta_{42}+\beta_{43}$
<13 years	***Male***	***80-84***	$\alpha_1+4\beta_{11}+\beta_{12}+\beta_{13}$	$\alpha_2+4\beta_{21}+\beta_{22}+\beta_{23}$	$\alpha_2+4\beta_{31}+\beta_{32}+\beta_{33}$	$\alpha_4+4\beta_{41}+\beta_{42}+\beta_{43}$
< 13 years	***Male***	***≥85***	$\alpha_1+5\beta_{11}+\beta_{12}+\beta_{13}$	$\alpha_2+5\beta_{21}+\beta_{22}+\beta_{23}$	$\alpha_2+5\beta_{31}+\beta_{32}+\beta_{33}$	$\alpha_4+5\beta_{41}+\beta_{42}+\beta_{43}$
<13 years	***Female***	***<70***	$\alpha_1+\beta_{11}-\beta_{12}+\beta_{13}$	$\alpha_2+\beta_{21}-\beta_{22}+\beta_{23}$	$\alpha_2+\beta_{31}-\beta_{32}+\beta_{33}$	$\alpha_4+\beta_{41}-\beta_{422}+\beta_{43}$
<13 years	***Female***	***70-75***	$\alpha_1+2\beta_{11}-\beta_{12}+\beta_{13}$	$\alpha_2+2\beta_{21}-\beta_{22}+\beta_{23}$	$\alpha_2+2\beta_{31}-\beta_{32}+\beta_{33}$	$\alpha_4+2\beta_{41}-\beta_{42}+\beta_{43}$
<13 years	***Female***	***75-80***	$\alpha_1+3\beta_{11}-\beta_{12}+\beta_{13}$	$\alpha_2+3\beta_{21}-\beta_{22}+\beta_{23}$	$\alpha_2+3\beta_{31}-\beta_{32}+\beta_{33}$	$\alpha_4+3\beta_{41}-\beta_{42}+\beta_{43}$
<13 years	***Female***	***80-84***	$\alpha_1+4\beta_{11}-\beta_{12}+\beta_{13}$	$\alpha_2+4\beta_{21}-\beta_{22}+\beta_{23}$	$\alpha_2+4\beta_{31}-\beta_{32}+\beta_{33}$	$\alpha_4+4\beta_{41}-\beta_{42}+\beta_{43}$
< 13 years	***Female***	***≥85***	$\alpha_1+5\beta_{11}-\beta_{12}+\beta_{13}$	$\alpha_2+5\beta_{21}-\beta_{22}+\beta_{23}$	$\alpha_2+5\beta_{31}-\beta_{32}+\beta_{33}$	$\alpha_4+5\beta_{41}-\beta_{42}+\beta_{43}$
≥13 years	***Male***	***<70***	$\alpha_1+\beta_{11}+\beta_{12}-\beta_{13}$	$\alpha_2+\beta_{21}+\beta_{22}-\beta_{23}$	$\alpha_2+\beta_{31}+\beta_{32}-\beta_{33}$	$\alpha_4+\beta_{41}+\beta_{42}-\beta_{43}$
≥13 years	***Male***	***70-75***	$\alpha_1+2\beta_{11}+\beta_{12}-\beta_{13}$	$\alpha_2+2\beta_{21}+\beta_{22}-\beta_{23}$	$\alpha_2+2\beta_{31}+\beta_{32}-\beta_{33}$	$\alpha_4+2\beta_{41}+\beta_{42}-\beta_{43}$
≥13 years	***Male***	***75-80***	$\alpha_1+3\beta_{11}+\beta_{12}-\beta_{13}$	$\alpha_2+3\beta_{21}+\beta_{22}-\beta_{23}$	$\alpha_2+3\beta_{31}+\beta_{32}-\beta_{33}$	$\alpha_4+3\beta_{41}+\beta_{42}-\beta_{43}$
≥13 years	***Male***	***80-84***	$\alpha_1+4\beta_{11}+\beta_{12}-\beta_{13}$	$\alpha_2+4\beta_{21}+\beta_{22}-\beta_{23}$	$\alpha_2+4\beta_{31}+\beta_{23}-\beta_{33}$	$\alpha_4+4\beta_{41}+\beta_{42}-\beta_{43}$
≥ 13 years	***Male***	***≥85***	$\alpha_1+5\beta_{11}+\beta_{12}-\beta_{13}$	$\alpha_2+5\beta_{21}+\beta_{22}-\beta_{23}$	$\alpha_2+5\beta_{31}+\beta_{23}-\beta_{33}$	$\alpha_4+5\beta_{41}+\beta_{42}-\beta_{43}$
≥13 years	***Female***	***<70***	$\alpha_1+\beta_{11}-\beta_{12}-\beta_{13}$	$\alpha_2+\beta_{21}-\beta_{22}-\beta_{23}$	$\alpha_2+\beta_{31}-\beta_{23}-\beta_{33}$	$\alpha_4+\beta_{41}-\beta_{42}-\beta_{43}$
≥13 years	***Female***	***70-75***	$\alpha_1+2\beta_{11}-\beta_{12}-\beta_{13}$	$\alpha_2+2\beta_{21}-\beta_{22}-\beta_{23}$	$\alpha_2+2\beta_{31}-\beta_{23}-\beta_{33}$	$\alpha_4+2\beta_{41}-\beta_{42}-\beta_{43}$
≥13 years	***Female***	***75-80***	$\alpha_1+3\beta_{11}-\beta_{12}-\beta_{13}$	$\alpha_2+3\beta_{21}-\beta_{22}-\beta_{23}$	$\alpha_2+3\beta_{31}-\beta_{23}-\beta_{33}$	$\alpha_4+3\beta_{41}-\beta_{42}-\beta_{43}$
≥13 years	***Female***	***80-84***	$\alpha_1+4\beta_{11}-\beta_{12}-\beta_{13}$	$\alpha_2+4\beta_{21}-\beta_{22}-\beta_{23}$	$\alpha_2+4\beta_{31}-\beta_{23}-\beta_{33}$	$\alpha_4+4\beta_{41}-\beta_{42}-\beta_{43}$
≥13 years	***Female***	***≥85***	$\alpha_1+5\beta_{11}-\beta_{12}-\beta_{13}$	$\alpha_2+5\beta_{21}-\beta_{22}-\beta_{23}$	$\alpha_2+5\beta_{31}-\beta_{23}-\beta_{33}$	$\alpha_4+5\beta_{41}-\beta_{42}-\beta_{43}$

Table IIb: Log odds for the cumulative model with age group (proportional odds), sex (proportional odds) and full-time education (non-proportional odds)

Full-time education (f)	Sex (g)	Age group (h)	$f_{fgh1}(\hat{\pi})$	$f_{fgh2}(\hat{\pi})$	$f_{fgh3}(\hat{\pi})$	$f_{fgh4}(\hat{\pi})$
<13 years	**Male**	**<70**	$\alpha_1+\beta_1+\beta_2+\beta_{13}$	$\alpha_2+\beta_1+\beta_2+\beta_{23}$	$\alpha_2+\beta_1+\beta_2+\beta_{33}$	$\alpha_4+\beta_1+\beta_2+\beta_{43}$
<13 years	**Male**	**70-75**	$\alpha_1+2\beta_1+\beta_2+\beta_{13}$	$\alpha_2+2\beta_1+\beta_2+\beta_{23}$	$\alpha_2+2\beta_1+\beta_2+\beta_{33}$	$\alpha_4+2\beta_1+\beta_2+\beta_{43}$
<13 years	**Male**	**75-80**	$\alpha_1+3\beta_1+\beta_2+\beta_{13}$	$\alpha_2+3\beta_1+\beta_2+\beta_{23}$	$\alpha_2+3\beta_1+\beta_2+\beta_{33}$	$\alpha_4+3\beta_1+\beta_2+\beta_{43}$
<13 years	**Male**	**80-84**	$\alpha_1+4\beta_1+\beta_2+\beta_{13}$	$\alpha_2+4\beta_1+\beta_2+\beta_{23}$	$\alpha_2+4\beta_1+\beta_2+\beta_{33}$	$\alpha_4+4\beta_1+\beta_2+\beta_{43}$
< 13 years	**Male**	**≥85**	$\alpha_1+5\beta_1+\beta_2+\beta_{13}$	$\alpha_2+5\beta_1+\beta_2+\beta_{23}$	$\alpha_2+5\beta_1+\beta_2+\beta_{33}$	$\alpha_4+5\beta_1+\beta_2+\beta_{43}$
<13 years	**Female**	**<70**	$\alpha_1+\beta_1-\beta_2+\beta_{13}$	$\alpha_2+\beta_1-\beta_2+\beta_{23}$	$\alpha_2+\beta_1-\beta_2+\beta_{33}$	$\alpha_4+\beta_1-\beta_2+\beta_{43}$
<13 years	**Female**	**70-75**	$\alpha_1+2\beta_1-\beta_2+\beta_{13}$	$\alpha_2+2\beta_1-\beta_2+\beta_{23}$	$\alpha_2+2\beta_1-\beta_2+\beta_{33}$	$\alpha_4+2\beta_1-\beta_2+\beta_{43}$
<13 years	**Female**	**75-80**	$\alpha_1+3\beta_1-\beta_2+\beta_{13}$	$\alpha_2+3\beta_1-\beta_2+\beta_{23}$	$\alpha_2+3\beta_1-\beta_2+\beta_{33}$	$\alpha_4+3\beta_1-\beta_2+\beta_{43}$
<13 years	**Female**	**80-84**	$\alpha_1+4\beta_1-\beta_2+\beta_{13}$	$\alpha_2+4\beta_1-\beta_2+\beta_{23}$	$\alpha_2+4\beta_1-\beta_2+\beta_{33}$	$\alpha_4+4\beta_1-\beta_2+\beta_{43}$
< 13 years	**Female**	**≥85**	$\alpha_1+5\beta_1-\beta_2+\beta_{13}$	$\alpha_2+5\beta_1-\beta_2+\beta_{23}$	$\alpha_2+5\beta_1-\beta_2+\beta_{33}$	$\alpha_4+5\beta_1-\beta_2+\beta_{43}$
≥13 years	**Male**	**<70**	$\alpha_1+\beta_1+\beta_2-\beta_{13}$	$\alpha_2+\beta_1+\beta_2-\beta_{23}$	$\alpha_2+\beta_1+\beta_2-\beta_{33}$	$\alpha_4+\beta_1+\beta_2-\beta_{43}$
≥13 years	**Male**	**70-75**	$\alpha_1+2\beta_1+\beta_2-\beta_{13}$	$\alpha_2+2\beta_1+\beta_2-\beta_{23}$	$\alpha_2+2\beta_1+\beta_2-\beta_{33}$	$\alpha_4+2\beta_1+\beta_2-\beta_{43}$
≥13 years	**Male**	**75-80**	$\alpha_1+3\beta_1+\beta_2-\beta_{13}$	$\alpha_2+3\beta_1+\beta_2-\beta_{23}$	$\alpha_2+3\beta_1+\beta_2-\beta_{33}$	$\alpha_4+3\beta_1+\beta_2-\beta_{43}$
≥13 years	**Male**	**80-84**	$\alpha_1+4\beta_1+\beta_2-\beta_{13}$	$\alpha_2+4\beta_1+\beta_2-\beta_{23}$	$\alpha_2+4\beta_1+\beta_2-\beta_{33}$	$\alpha_4+4\beta_1+\beta_2-\beta_{43}$
≥ 13 years	**Male**	**≥85**	$\alpha_1+5\beta_1+\beta_2-\beta_{13}$	$\alpha_2+5\beta_1+\beta_2-\beta_{23}$	$\alpha_2+5\beta_1+\beta_2-\beta_{33}$	$\alpha_4+5\beta_1+\beta_2-\beta_{43}$
≥13 years	**Female**	**<70**	$\alpha_1+\beta_1-\beta_2-\beta_{13}$	$\alpha_2+\beta_1-\beta_2-\beta_{23}$	$\alpha_2+\beta_1-\beta_2-\beta_{33}$	$\alpha_4+\beta_1-\beta_2-\beta_{43}$
≥13 years	**Female**	**70-75**	$\alpha_1+2\beta_1-\beta_2-\beta_{13}$	$\alpha_2+2\beta_1-\beta_2-\beta_{23}$	$\alpha_2+2\beta_1-\beta_2-\beta_{33}$	$\alpha_4+2\beta_1-\beta_2-\beta_{43}$
≥13 years	**Female**	**75-80**	$\alpha_1+3\beta_1-\beta_2-\beta_{13}$	$\alpha_2+3\beta_1-\beta_2-\beta_{23}$	$\alpha_2+3\beta_1-\beta_2-\beta_{33}$	$\alpha_4+3\beta_1-\beta_2-\beta_{43}$
≥13 years	**Female**	**80-84**	$\alpha_1+4\beta_1-\beta_2-\beta_{13}$	$\alpha_2+4\beta_1-\beta_2-\beta_{23}$	$\alpha_2+4\beta_1-\beta_2-\beta_{33}$	$\alpha_4+4\beta_1-\beta_2-\beta_{43}$
≥13 years	**Female**	**≥85**	$\alpha_1+5\beta_1-\beta_2-\beta_{13}$	$\alpha_2+5\beta_1-\beta_2-\beta_{23}$	$\alpha_2+5\beta_1-\beta_2-\beta_{33}$	$\alpha_4+5\beta_1-\beta_2-\beta_{43}$

Table IIc: Log odds for the unconstrained partial proportional odds model with the 1st order interaction terms using the Townsend disability score

Full-time education (f)	Sex (g)	Age group (h)	$f_{fgh1}(\hat{\pi})$	$f_{fgh2}(\hat{\pi})$	$f_{fgh3}(\hat{\pi})$	$f_{fgh4}(\hat{\pi})$
<13 years	Male	<70	$\alpha_1+\beta_1+\beta_2+(\beta_3+\gamma_{13})+\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+\beta_1+\beta_2+(\beta_3+\gamma_{23})+\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+\beta_1+\beta_2+(\beta_3+\gamma_{33})+\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+\beta_1+\beta_2+(\beta_3+\gamma_{43})+\beta_4+(\beta_6+\gamma_{46})$
<13 years	Male	70-75	$\alpha_1+2\beta_1+\beta_2+(\beta_3+\gamma_{13})+2\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+2\beta_1+\beta_2+(\beta_3+\gamma_{23})+2\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+2\beta_1+\beta_2+(\beta_3+\gamma_{33})+2\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+2\beta_1+\beta_2+(\beta_3+\gamma_{43})+2\beta_4+(\beta_6+\gamma_{46})$
<13 years	Male	75-80	$\alpha_1+3\beta_1+\beta_2+(\beta_3+\gamma_{13})+3\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+3\beta_1+\beta_2+(\beta_3+\gamma_{23})+3\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+3\beta_1+\beta_2+(\beta_3+\gamma_{33})+3\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+3\beta_1+\beta_2+(\beta_3+\gamma_{43})+3\beta_4+(\beta_6+\gamma_{46})$
<13 years	Male	80-84	$\alpha_1+4\beta_1+\beta_2+(\beta_3+\gamma_{13})+4\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+4\beta_1+\beta_2+(\beta_3+\gamma_{23})+4\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+4\beta_1+\beta_2+(\beta_3+\gamma_{33})+4\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+4\beta_1+\beta_2+(\beta_3+\gamma_{43})+4\beta_4+(\beta_6+\gamma_{46})$
< 13 years	Male	≥85	$\alpha_1+5\beta_1+\beta_2+(\beta_3+\gamma_{13})+5\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+5\beta_1+\beta_2+(\beta_3+\gamma_{23})+5\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+5\beta_1+\beta_2+(\beta_3+\gamma_{33})+5\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+5\beta_1+\beta_2+(\beta_3+\gamma_{43})+5\beta_4+(\beta_6+\gamma_{46})$
<13 years	Female	<70	$\alpha_1+\beta_1-\beta_2+(\beta_3+\gamma_{13})-\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+\beta_1-\beta_2+(\beta_3+\gamma_{23})-\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+\beta_1-\beta_2+(\beta_3+\gamma_{33})-\beta_4-(\beta_6+\gamma_{36})$	$\alpha_4+\beta_1-\beta_2+(\beta_3+\gamma_{43})-\beta_4-(\beta_6+\gamma_{46})$
<13 years	Female	70-75	$\alpha_1+2\beta_1-\beta_2+(\beta_3+\gamma_{13})-2\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+2\beta_1-\beta_2+(\beta_3+\gamma_{23})-2\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+2\beta_1-\beta_2+(\beta_3+\gamma_{33})-2\beta_4-(\beta_6+\gamma_{36})$	$\alpha_4+2\beta_1-\beta_2+(\beta_3+\gamma_{43})-2\beta_4-(\beta_6+\gamma_{46})$
<13 years	Female	75-80	$\alpha_1+3\beta_1-\beta_2+(\beta_3+\gamma_{13})-3\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+3\beta_1-\beta_2+(\beta_3+\gamma_{23})-3\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+3\beta_1-\beta_2+(\beta_3+\gamma_{33})-3\beta_4-(\beta_6+\gamma_{36})$	$\alpha_4+3\beta_1-\beta_2+(\beta_3+\gamma_{43})-3\beta_4-(\beta_6+\gamma_{46})$
<13 years	Female	80-84	$\alpha_1+4\beta_1-\beta_2+(\beta_3+\gamma_{13})-4\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+4\beta_1-\beta_2+(\beta_3+\gamma_{23})-4\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+4\beta_1-\beta_2+(\beta_3+\gamma_{33})-4\beta_4-(\beta_6+\gamma_{36})$	$\alpha_4+4\beta_1-\beta_2+(\beta_3+\gamma_{43})-4\beta_4-(\beta_6+\gamma_{46})$
< 13 years	Female	≥85	$\alpha_1+5\beta_1-\beta_2+(\beta_3+\gamma_{13})-5\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+5\beta_1-\beta_2+(\beta_3+\gamma_{23})-5\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+5\beta_1-\beta_2+(\beta_3+\gamma_{33})-5\beta_4-(\beta_6+\gamma_{36})$	$\alpha_4+5\beta_1-\beta_2+(\beta_3+\gamma_{43})-5\beta_4-(\beta_6+\gamma_{46})$
≥13 years	Male	<70	$\alpha_1+\beta_1+\beta_2-(\beta_3+\gamma_{13})+\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+\beta_1+\beta_2-(\beta_3+\gamma_{23})+\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+\beta_1+\beta_2-(\beta_3+\gamma_{33})+\beta_4-(\beta_6+\gamma_{36})$	$\alpha_4+\beta_1+\beta_2-(\beta_3+\gamma_{43})+\beta_4-(\beta_6+\gamma_{46})$
≥13 years	Male	70-75	$\alpha_1+2\beta_1+\beta_2-(\beta_3+\gamma_{13})+2\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+2\beta_1+\beta_2-(\beta_3+\gamma_{23})+2\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+2\beta_1+\beta_2-(\beta_3+\gamma_{33})+2\beta_4-(\beta_6+\gamma_{36})$	$\alpha_4+2\beta_1+\beta_2-(\beta_3+\gamma_{43})+2\beta_4-(\beta_6+\gamma_{46})$
≥13 years	Male	75-80	$\alpha_1+3\beta_1+\beta_2-(\beta_3+\gamma_{13})+3\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+3\beta_1+\beta_2-(\beta_3+\gamma_{23})+3\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+3\beta_1+\beta_2-(\beta_3+\gamma_{33})+3\beta_4-(\beta_6+\gamma_{36})$	$\alpha_4+3\beta_1+\beta_2-(\beta_3+\gamma_{43})+3\beta_4-(\beta_6+\gamma_{46})$
≥13 years	Male	80-84	$\alpha_1+4\beta_1+\beta_2-(\beta_3+\gamma_{13})+4\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+4\beta_1+\beta_2-(\beta_3+\gamma_{23})+4\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+4\beta_1+\beta_2-(\beta_3+\gamma_{13})+4\beta_4-(\beta_6+\gamma_{16})$	$\alpha_4+4\beta_1+\beta_2-(\beta_3+\gamma_{43})+4\beta_4-(\beta_6+\gamma_{46})$
≥ 13 years	Male	≥85	$\alpha_1+5\beta_1+\beta_2-(\beta_3+\gamma_{13})+5\beta_4-(\beta_6+\gamma_{16})$	$\alpha_2+5\beta_1+\beta_2-(\beta_3+\gamma_{23})+5\beta_4-(\beta_6+\gamma_{26})$	$\alpha_3+5\beta_1+\beta_2-(\beta_3+\gamma_{33})+5\beta_4-(\beta_6+\gamma_{36})$	$\alpha_4+5\beta_1+\beta_2-(\beta_3+\gamma_{43})+5\beta_4-(\beta_6+\gamma_{46})$
≥13 years	Female	<70	$\alpha_1+\beta_1-\beta_2-(\beta_3+\gamma_{13})-\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+\beta_1-\beta_2-(\beta_3+\gamma_{23})-\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+\beta_1-\beta_2-(\beta_3+\gamma_{33})-\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+\beta_1-\beta_2-(\beta_3+\gamma_{43})-\beta_4+(\beta_6+\gamma_{46})$
≥13 years	Female	70-75	$\alpha_1+2\beta_1-\beta_2-(\beta_3+\gamma_{13})-2\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+2\beta_1-\beta_2-(\beta_3+\gamma_{23})-2\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+2\beta_1-\beta_2-(\beta_3+\gamma_{33})-2\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+2\beta_1-\beta_2-(\beta_3+\gamma_{43})-2\beta_4+(\beta_6+\gamma_{46})$
≥13 years	Female	75-80	$\alpha_1+3\beta_1-\beta_2-(\beta_3+\gamma_{13})-3\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+3\beta_1-\beta_2-(\beta_3+\gamma_{23})-3\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+3\beta_1-\beta_2-(\beta_3+\gamma_{33})-3\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+3\beta_1-\beta_2-(\beta_3+\gamma_{43})-3\beta_4+(\beta_6+\gamma_{46})$
≥13 years	Female	80-84	$\alpha_1+4\beta_1-\beta_2-(\beta_3+\gamma_{13})-4\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+4\beta_1-\beta_2-(\beta_3+\gamma_{23})-4\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+4\beta_1-\beta_2-(\beta_3+\gamma_{33})-4\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+4\beta_1-\beta_2-(\beta_3+\gamma_{43})-4\beta_4+(\beta_6+\gamma_{46})$
≥13 years	Female	≥85	$\alpha_1+5\beta_1-\beta_2-(\beta_3+\gamma_{13})-5\beta_4+(\beta_6+\gamma_{16})$	$\alpha_2+5\beta_1-\beta_2-(\beta_3+\gamma_{23})-5\beta_4+(\beta_6+\gamma_{26})$	$\alpha_3+5\beta_1-\beta_2-(\beta_3+\gamma_{33})-5\beta_4+(\beta_6+\gamma_{36})$	$\alpha_4+5\beta_1-\beta_2-(\beta_3+\gamma_{43})-5\beta_4+(\beta_6+\gamma_{46})$

APPENDIX III - LIPSITZ'S (1996) GOODNESS-OF-FIT STATISTICS AND THE RESULTS

Section 1: Lipsitz's (1996) statistics: The generalisation of the Hosmer-Lemeshow goodness-of-fit statistic using ordinal response data

Let the general form of an ordinal regression model be specified (as in (3.5)):

$$F(\pi) = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_{jk} \tag{A18}$$

To form the goodness-of-fit statistic, as proposed by Lipsitz (1996), one assigns a score s_j to response category j. The assigned scores may in some instance be the actual numerical response or the mid-point of the interval when the response is a crude grouping of an underlying continuous variable. Then one can obtain the 'fitted' score or predicted mean score. Suppose that s_j is the value assigned to category j with $s=(s_1,\ldots s_c)'$ then the observed score for the i^{th} group is:

$$Z_i = \sum_{j=1}^{c} s_j Y_{ij} = s'Y_i. \tag{A19}$$

The mean score is

$$\mu_i = \mu_i(\beta) = E(Z_i) = \sum_{j=1}^{c} s_j \hat{\pi}_{ij} = s'\hat{\pi}_i \tag{A20}$$

and the predicted mean sore is

$$\hat{\mu}_i = \sum_{j=1}^{c} s_j \hat{p}_{ij} \tag{A21}$$

To form the goodness-of-fit statistic one then groups or partitions subjects into regions based on the predicted mean scores $\hat{\mu}_i$. Following the approach of Hosmer-Lemeshow (1980) with binary data, 10 groups of approximately equal size are formed. The first group contains the N/10 subjects with the smallest predicted mean score, and the last group contains the N/10 subjects with the largest predicted mean score. Given the partition of the data, the goodness-of-fit statistic is formulated by defining the G-1 group indicators,

$$I_{ig} = \begin{cases} 1 \\ 0 \end{cases}, \text{ 1 if } \mu_i \text{ is in region } g, \text{ and 0 otherwise,} \qquad \text{(A22)}$$

g=1,.. G-1. Then to assess the goodness-of-fit of the ordinal model, one considers the alternative to model (A18) namely

$$F(\pi) = \alpha_j + \sum_{k=1}^{p} x_{ik}\beta_{jk} + \sum_{g=1}^{G-1} I_{ig}\gamma_g \,. \qquad \text{(A23)}$$

If model (A18) is correctly specified then $\gamma_1 = \gamma_2 = ...\gamma_{G-1} = 0$ in equation (A23), regardless of how the regions are chosen and regardless of which scores are used.

Section 2: Results from the model fitting procedure

Health Status data

(i) Linear Regression Model

The linear regression model was fitted using the methods outlined in section 5.2.1. Both covariates were fitted as categorical variables and these were found to be significant in the model, when taken individually or added in together. The interaction 'heart attack x 'smoke' was found not to be significant (*t-test* = -0.16; p =0.87 on 1 df) and therefore only the main effects model was considered further.

(ii) Binary Logistic Regression Models

The methods used to fit the binary logistic regression models are detailed in section 5.2.2. The deviance which was based on the change in the –2log-likelihood of the main effects and saturated model provided a test for the significance of the interaction term. The 1st order interaction term for the binary logistic regression model based on the first cut-point was found to be significant (deviance =4.41 on 1-df), whilst the 1st order interaction terms for the binary models based on the other cut-points were non-significant (cut-point 2: deviance =0.16 on 1-df and cut-point 3: deviance =0.004 on 1-df).

(iii) Polytomous Model

The method used to fit the polytomous model is outlined in section 5.2.3.1. The change in the -2log-likelihoods for the saturated and main effects models (as expressed in matrices (5.2) and (5.3) respectively) was not significant, implying a non-significant interaction term (change in –2log-likelihood=6.37 on 3-df.:-2log-likelihood for the saturated model was 29322.79 on 12-df and the main effects model was 29329.16 on 9-df). Therefore only the mains effect model was considered further.

(iv) Different slope Cumulative Logit Model

The different slopes cumulative logit model (5.4) was fitted using the main effects. The interaction term was not considered here, as the purpose of this model was to visually assess how the main effects were behaving with respect to the cut-points.

(v) Proportional odds Model

Both main effects were fitted with the assumption of proportional odds (as given in (5.5)). The interaction term was also assumed to have proportional odds (as given in (5.7)). The interaction term was not found to be significant (χ^2_1=0.01; p =0.94) and therefore only the main effect proportional odds model was considered further.

(vi) Unconstrained Partial Proportional Odds Model

Initially the model with the main effects and the 1st order interaction term was fitted (as in (5.13)). This model had 12 parameters to estimate (equal to the number of logits in the contingency table). Some of these parameters were constrained to have equal slopes and therefore the model was not a saturated one. There was evidence that the effects that made up the interaction term were not significant (H_{01}: $\beta_1\beta_2$=0: χ^2_1=0.02; p =0.90; H_{02}: $\beta_2\gamma_{21}$=$\beta_2\gamma_{31}$=0; χ^2_2=4.50; p =0.11). Thus the interaction model was not considered any further and the results were based on the main effects model.

(vii) Constrained Partial Proportional Odds Model

The constrained partial proportional odds model with the 1st order interaction term was fitted (as in (5.22)) and there was no evidence of a significant interaction term (H_{01}: $\beta_1\beta_2$=0:

χ^2_1=0.09 p =0.76; H_{02}: $\beta_2\gamma_1$=0: χ^2_1=0.19 p =0.66). Again, only the main effects model was considered further.

(viii) Constant slope Adjacent Category Model

The interaction term of the constant slope adjacent category model (as fitted using methods in section 5.2.3.3 (i)) was found not to be significant (χ_1^2=0.37; p =0.54). This interaction term was also assumed to have constant slopes.

(ix) Different slopes Adjacent Category Model

The adjacent category model was fitted with the different parameter estimates over the cut-points (as given in section 5.2.3.3 (ii)). The 1st order interaction term was found not to be significant (χ_3^2=6.68; p =0.08) for this model. The main effects model was taken as the final model.

(x) Fully constrained Continuation Ratio Model

The interaction term was fitted in a similar way to the proportional odds model. This term was not significant (χ_1^2=0.35; p =0.56).

(xi) Continuation Ratio Models

Binary Analysis: The 1st order interaction terms for the binary logistic regression models based on the second and third cut-points were found not to be significant (cut-point 2 model: χ_1^2=0.35 with p =0.55 and model based on cut-point 3: χ_1^2=0.02 with p =0.90). However, the first order interaction term was significant for the model based on cut-point 1 (χ_1^2=6.64 with p =0.01). To provide an overall conclusion for the continuation ratio model there was a need to keep all the models consistent with regards to the parameters fitted. Thus the models based on cut-points 2 and 3 were each fitted with an interaction term.

Unconstrained continuation ratio model: The unconstrained continuation ratio model with the 1st order interaction term was fitted using the specification as detailed in section 5.2.3.4 (d). The first order interaction term was not significant (χ_3^2=6.72 with p =0.08). Thus the main effects model was taken as the final model.

(xii) Stereotype Model

The main effects stereotype model was fitted initially (using the methods described in section 5.2.3.5) and the first order interaction term was subsequently added in. To assess whether this latter term was significant or not, bootstrapping was used. The observed change in the –2log-likelihood given the main and saturated model was 0.037 (main effect model –2log-likelihood=29343.73 on fitting 5 parameters and model with 1st order interaction term –2log-likelihood=29343.69 with fitting 6 parameters). There were a 100 bootstrap samples and for each one a saturated and main effects model was fitted and the change in the –2log-likelihood computed. There were 100 change values that formed a distribution. Assessing the distribution, 14% of the change values from the bootstrapped data were below the observed change value and 86% were above. Thus the ASL was 0.86 and this suggested that the null hypothesis could not be rejected and the interaction term was therefore not significant. The main effects model was used to summarise the results.

Townsend Disability Scale data

Forward selection procedure was used to include the terms in the models.

(i) Linear Regression Model

The statistical method used to fit the linear regression model is given in section 5.3.1. All three covariates when fitted individually were found to be significant. The 'sex' and 'age group ' covariates were fitted together in the second stage of modelling and then 'full-time education' was added in. All the main effects when fitted together were found to be statistically significant. The first order interaction terms were individually added into the main effects model and it was found that only 'age group *x* sex' was significant (*t-test statistic*=39.69; p <0.0001) with the other two interactions non-significant: 'sex *x* full-time education' (*t-test statistic* =-1.23; p =0.22 on 1 df) and 'age group *x* full-time education' (*t-test statistic* =-1.75; p =0.08 on 1 df).

(ii) Binary Logistic Regression Model

The binary logistic model (as outlined in section 5.3.2) with all three main effects was initially taken and each first-order interaction term was added in. No interaction terms were found significant ('age-group *x* sex': χ^2_4= 5.20; p =0.27; 'sex *x* full-time education': χ^2_1= 0.01;

p =0.93; 'age group x full-time education': χ^2_4= 2.21; p =0.70). Therefore the main effect model was considered as the final model.

(iii) Polytomous Models

The polytomous model was fitted as detailed in section 5.3.3.1. Each 1st order interaction term was added individually to the main effects model and only 'age group x sex' interaction term was found to be significant (χ^2_4=29.16; p <0.0001: df =4 since there were four cut-point specific effects which made up the interaction term). The other interaction terms were not statistically significant ('age group x full-time education': χ^2_4=3.66; p =0.45; 'sex x full-time education': χ^2_4=2.53; p =0.64).

(iv) Proportional odds/ Different slopes Cumulative Logit Model

The 1st order interaction term was not fitted as interest was only focused on the behaviour of the main covariates. However, the proportional odds model was fitted with interaction terms. It was found that 'sex x age group' was the only significant interaction term ('sex x age group': χ^2_1=7.34; p =0.007; 'age group x full-time education': χ^2_1= 0.33; p =0.56; 'sex x full-time education': χ^2_1=0.82; p =0.36).

(v) Unconstrained Partial Proportional Odds Model

The starting point for the unconstrained partial proportional odds model was model (5.31). Each interaction term model as detailed in (5.32), (5.33) and (5.34) was fitted individually into the model and it was found that the 'age group x sex' term was significant (test based on H_{01}: β_4=0: χ^2_1=7.61; p =0.01). The interaction 'age group x full-time education' was not significant (test based on H_{02}: β_5=0: χ^2_1=0.53; p =0.68 and H_{03}: $\gamma_{25}=\gamma_{35}=\gamma_{45}$=0: χ^2_3=5.55; p =0.14). The interaction 'sex x full-time education' was significant (test based on H_{04}: β_6=0: χ^2_1=3.10; p =0.01 and H_{05}: $\gamma_{26}=\gamma_{36}=\gamma_{46}$=0: χ^2_3=9.31; p =0.02). The 'sex x full-time education' term was added into the 1st order interaction model (5.32) and both interaction terms remained significant. The final model could be represented using model (5.34) where both 'age group x sex' and 'sex x full-time education' interaction terms were included in the main effects model.

(vi) Constant slope Adjacent Category Model

The 1st order interaction terms were tested for the latter model (as described in section 5.3.3.3 (i)) where all the main effects were assumed to have a constant slope over the cut-points. There was evidence that none of these terms were significant and therefore the odds ratios were based on the main effect model ('age group *x* sex': χ^2_1= 1.33; p =0.29; 'age group *x* full-time education': χ^2_1= 0.16; p =0.69; 'sex *x* full-time education': χ^2_1= 0.28; p =0.60).

(vii) Different slopes Adjacent Category Model

The different slopes adjacent category model was fitted using methods specified in section 5.3.3.3 (ii). Each first-order interaction term was added individually into the main effects model. It was found that 'age group *x* sex' was significant (χ^2_4= 25.73; $p < 0.0001$) with the other two first order interaction terms found to be non-significant ('sex *x* full-time education': χ^2_4= 2.50; p =0.64; 'age group *x* full-time education': χ^2_4= 3.32; p =0.51).

(viii) Fully constrained Continuation Ratio Model

The first order interaction terms were tested for the constant slope models and it was found that none of the interaction terms was significant ('age group *x* sex': χ^2_1= 0.60; p =0.44; 'age group *x* full-time education': χ^2_1= 0.33; p =0.56; 'sex *x* full-time education': χ^2_1= 0.58; p =0.45). Therefore the analysis was based on the main effects model.

(ix) Different slopes Continuation Ratio Models

(a) Using separate binary logistic regression models

Separate binary logistic regression models were fitted (using methods given in section 5.3.3.4 (ii)) to each cut-point based on the continuation ratio logits. The overall results from these models were corrected for multiple testing. Since four binary logistic regression models were fitted, the Type I error rate was α=0.05/4=0.01.

Model for cut-point 1: *'none'* v. (*'slight'*, *'some'*, *'appreciable'*, *'severe+ v. severe'*)

For the model based on cut-point 1 the main effects were fitted and then each first order interaction term was added into the model. Of the three first-order interaction terms, only 'age group *x* sex' was found to be significant (χ_1^2=8.90 with p =0.003) and the other interaction terms were non-significant ('age group *x* full-time education': χ_1^2=0.18 with p =0.67; 'sex *x* full-time education': χ_1^2=0.94 with p =0.33).

<u>Model for cut-point 2: '*slight*' v. ('*some*', '*appreciable*', '*severe+ v. severe*')</u>

The main effects were again fitted with each first order interaction term and again only the interaction of 'age group *x* sex' was found significant (χ_1^2=11.91 with p =0.0006). The other first order interaction terms were non-significant ('age group *x* full-time education': χ_1^2=0.03 with p =0.87; 'sex *x* full-time education': χ_1^2=0.76 with p =0.38).

<u>Model for cut-point 3: '*some*' v. ('*appreciable*', '*severe+ v. severe*')</u>

Each first order interaction term was added to the main effect model. After allowing for multiple testing, there was indication that 'age group *x* sex' term was significant but the other interaction terms were not ('age group *x* sex': χ_1^2=5.89 with p =0.002; 'age group *x* full-time education': χ_1^2=1.07 with p =0.30; 'sex *x* full-time education': χ_1^2=0.99 with p =0.32).

<u>Model 4: '*appreciable*' v. '*severe+ v. severe*'</u>

The interaction 'age group *x* sex' was significant (χ_1^2=13.40 with p =0.0003), and the other interaction terms were non-significant ('age group *x* full-time education': χ_1^2=0.32 with p =0.57; 'sex *x* full-time education': χ_1^2=0.02 with p =0.88).

<u>*(b) Using the unconstrained continuation ratio model*</u>

The main effects unconstrained continuation ratio model was fitted and each of the three interaction terms were added in individually. There was evidence that the 'age group *x* sex' interaction term was significant (χ_4^2=36.14 with p <0.0001). The other two 1st order interaction terms were non-significant ('sex *x* full-time education': χ_4^2=2.51 with p =0.64; 'age-group *x* full-time education': χ_4^2=1.56 with p =0.82).

(x) Stereotype Model

Each individual 1st order interaction term was added to the main effects one-dimensional stereotype model. The observed change in the –2log-likelihood of the main effects model and the model with the interaction term- 'age group *x* sex' was 4.98 (where the –2log-likelihood for the main effect model was 34459.79 fitting 7 parameters and the –2log-likelihood for the interaction model was 34454.806 fitting 8 parameters). In the distribution of change values based on the bootstrapped samples, there were 50.4% values less than the observed change and 49.5 % that were above. The ASL was 0.495 and thus the null hypothesis was rejected, as there was borderline evidence that the two models were different. As a result, the interaction 'age group *x* sex' was an additional term included in the main effects model.

The observed change in the –2log-likelihood of the main effects model and the model with the interaction of 'sex x full-time education' was 0.31 (where -2log-likelihood for the model with interaction of sex *x* full-time education=34459.48 fitting 8 parameters). The main and interaction models were fitted to the 100 bootstrapped samples and the distribution of the change in the –2log-likelihoods formed. There was evidence that 32.6% of the change values were less than the observed change value and 67.3% were above it. As the ASL was 0.67, the null hypothesis could not be rejected and therefore the interaction 'sex *x* full-time education' was found not to be significant.

The observed change in the –2log-likelihoods of the main effects model and the model with the interaction of 'age group *x* full-time education' was 0.01 (where –2log-likelihood for model with interaction age group *x* full-time education =34459.78 fitting 8 parameters). The position of this observed change was again assessed in the distribution of the 100 change values in the –2log-likelihood values between the two models. There was indication that 8.9% of the values were below the observed change values and 91.1% were above it (i.e. ASL =0.91), indicating that one could not reject the null hypothesis, and therefore the latter interaction term was not significant.

Table IIIa: Parameter estimates and their standard errors for the individual binary logistic regression models based on the Polytomous Model for the Townsend disability score

Variable	Parameters	Cut-point 1*	Cut-point 2*	Cut-point 3*	Cut-point 4*
Intercept(α_j)	Parameter estimate	4.5765	3.2721	2.3507	0.6235
	s.e	0.1280	0.1272	0.1253	0.1400
Age group	Parameter estimate	-1.0551	-0.7192	-0.4424	-0.1714
	s.e.	0.0333	0.0323	0.0308	0.0334
Sex	Parameter estimate	0.1118	-0.1877	-0.2920	-0.4089
	s.e.	0.1087	0.1089	0.1080	0.1212
f-t education	Parameter estimate	-0.4362	-0.3809	-0.3267	-0.1090
	s.e.	0.0727	0.0712	0.0687	0.0801
Age group x	Parameter estimate	0.1318	0.1485	0.1261	0.1225
sex	s.e.	0.0333	0.0322	0.0307	0.0335
Hosmer-	df	7	7	7	7
Lemeshow	Test-statistic	59.2038	49.7424	37.996	9.8199
statistic	*p*	<0.0001	<0.0001	<0.0001	0.1324

** - Cut-point 1 refers to 'none' v. 'severe+ v. severe'; Cut-point 2 refers to 'slight' v. 'severe + v. severe'; Cut-point 3 refers to 'some' v. 'severe + v. severe' and cut-point 4 refers to 'appreciable' v. 'severe + v. severe'.*

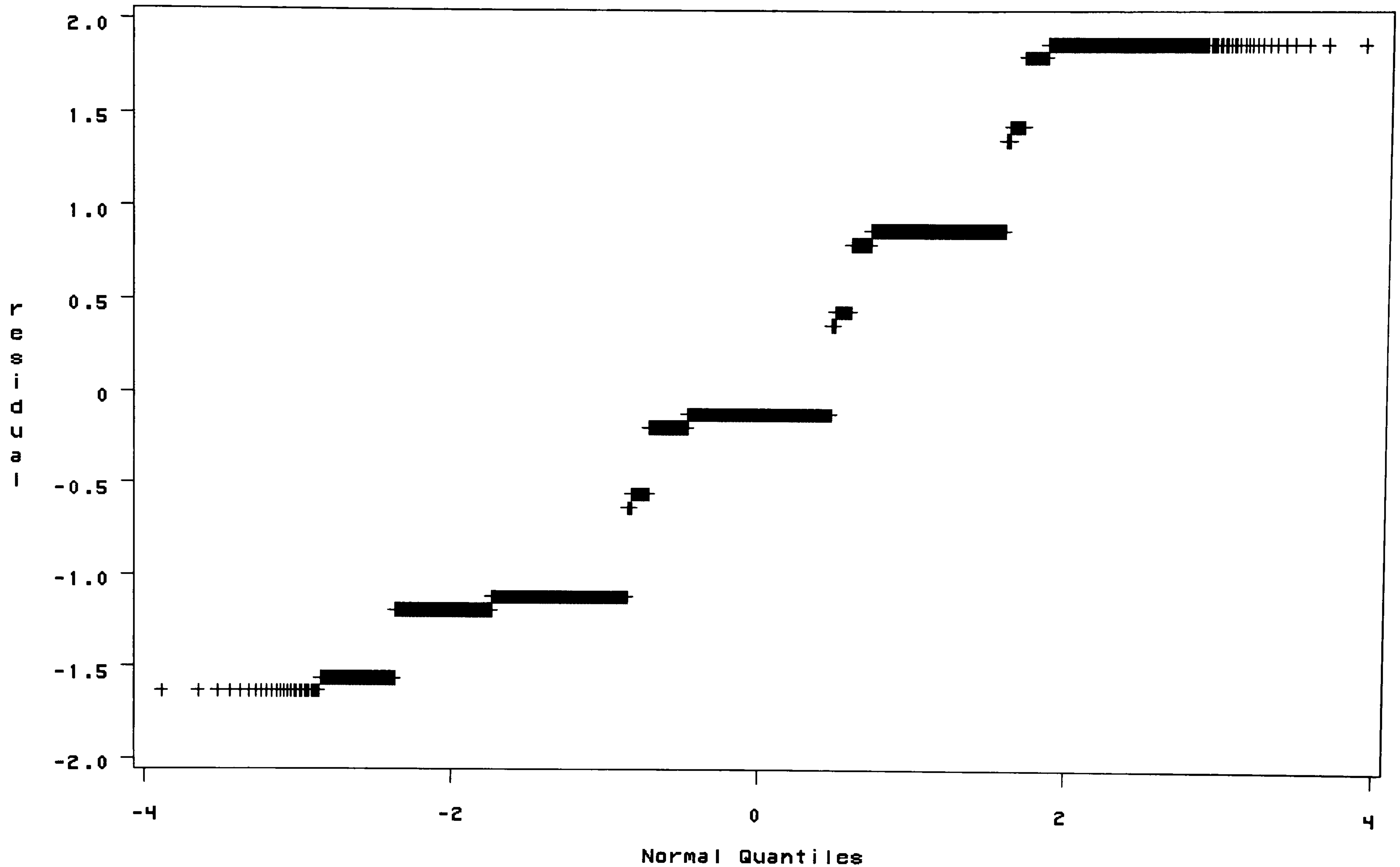
Table IIIb: Parameter estimates and their standard errors for the individual binary logistic regression models based on the Different slopes Adjacent Category Model for the Townsend disability score

Variable	Parameters	Cut-point 1	Cut-point 2	Cut-point 3	Cut-point 4
Age group	Parameter estimate	-0.3565	-0.2836	-0.2540	-0.1714
	s.e.	0.0226	0.0242	0.0302	0.0334
sex	Parameter estimate	0.2726	0.0524	0.1459	-0.4089
	s.e.	0.0572	0.0711	0.1023	0.1212
f-t education	Parameter estimate	-0.0782	-0.0713	-0.2104	-0.1090
	s.e.	0.0395	0.0467	0.0653	0.0801
age group x	Parameter estimate	-0.00577	0.0391	-0.00353	0.1225
sex	s.e.	0.02665	0.0242	0.0302	0.0335
Hosmer-	df	6	8	8	6
Lemeshow	Test-statistic	1.8574	9.2207	26.8266	9.8199
statistic	*p*	0.9323	0.3240	0.0004	0.1342

** - Cut-point 1 refers to 'none' v. 'slight'; Cut-point 2 refers to 'slight' v. 'some'; Cut-point 3 refers to 'some' v. 'appreciable' and cut-point 4 refers to 'appreciable' v. 'severe + v. severe'.*

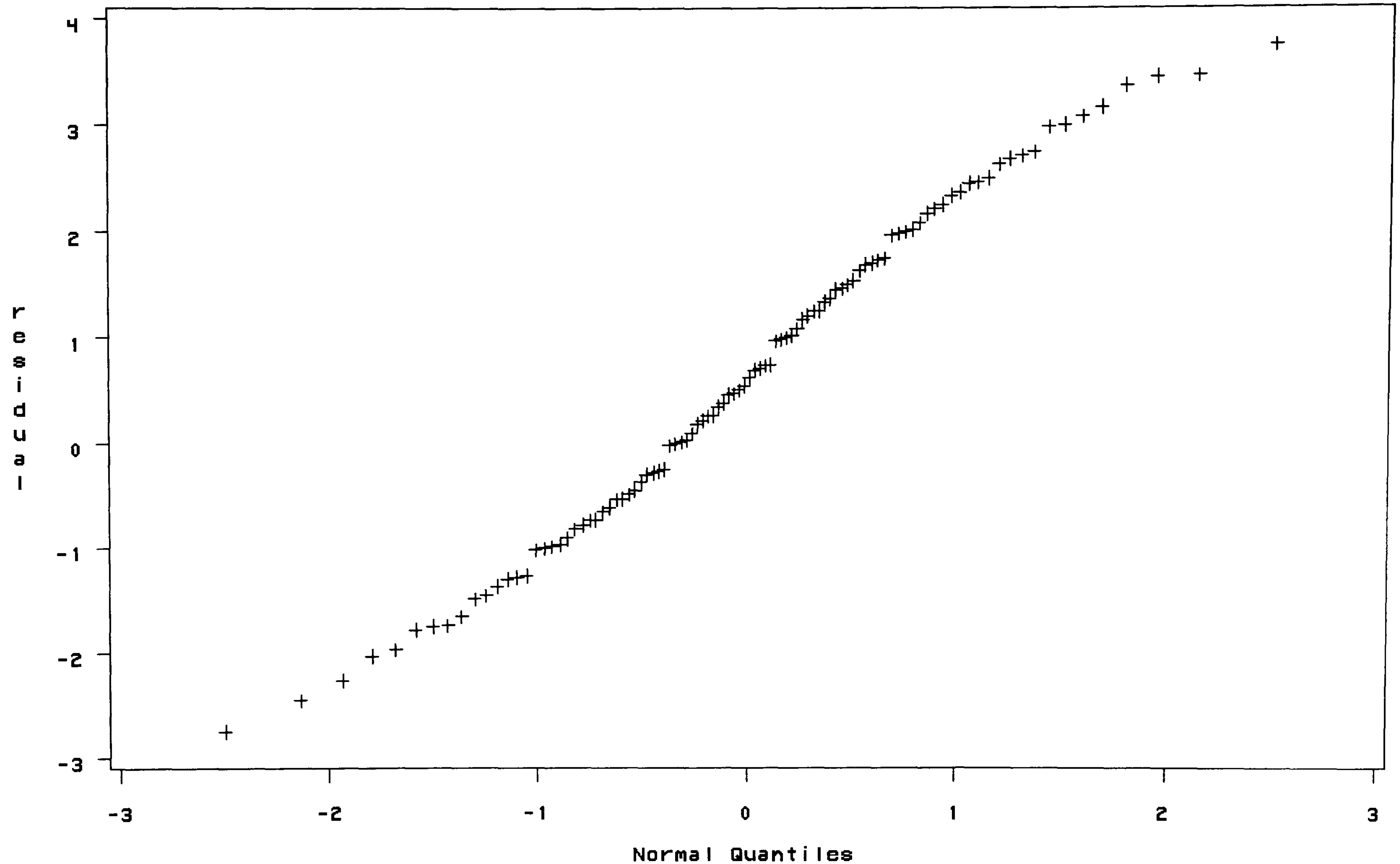
Plot IIIa

Normal Plot for the Standard Normal deviate against the residuals
for the Health Status Data for the main effects model



Plot IIIb

Normal Plot of the Standard normal deviate against the residuals
for the Townsend Disability Score for the main effects and interaction of age and sex model



APPENDIX IV: PROOF OF THE CONSTANT ODDS OF THE PROPORTIONAL ODDS MODEL AND THE MONOTONICALLY INCREASING ODDS OF THE POLYTOMOUS MODEL

From the data in this thesis, there was evidence that the polytomous model has monotonically increasing odds ratios. One can prove that if a constant odds assumption holds for the proportional odds model, then the odds for the polytomous model will always be monotonically increasing.

Given ordinal outcomes, and without loss of generality let us assume that there are 3 ordered *y*-response categories and one covariate on two levels.

Table IVa: Ordinal response with three categories and one covariate on two levels

	Outcome		
Covariate			
	y_1	y_2	y_3
x_1	π_1	π_2	π_3
x_2	φ_1	φ_2	φ_3

From this $\pi_1 + \pi_2 + \pi_3 = 1$ and $\varphi_1 + \varphi_2 + \varphi_3 = 1$. Also it has to be assumed that $\pi_1 > 0$, $\pi_2 > 0$, $\pi_3 > 0$ and also $\varphi_1 > 0$, $\varphi_2 > 0$ and $\varphi_3 > 0$.

Assuming that the constant odds ratio as obtained by the cumulative odds model is denoted by λ then it can be written

$$\frac{\pi_1}{1-\pi_1} = \lambda \frac{\varphi_1}{1-\varphi_1} \qquad \text{(A24)}$$

and

$$\frac{\pi_1 + \pi_2}{1-(\pi_1 + \pi_2)} = \lambda \left(\frac{\varphi_1 + \varphi_2}{1-(\varphi_1 + \varphi_2)} \right) \qquad \text{(A25)}$$

Assume that $\lambda > 0$ and also π_1, π_2 and λ are given. We wish to prove that the polytomous odds ratios are monotonic. Taking the y_1 as the referent category we wish to prove

$\frac{\pi_2/\pi_1}{\varphi_2/\varphi_1} > \frac{(1-(\pi_1 + \pi_2))/\pi_1}{(1-(\varphi_1 + \varphi_2))/\varphi_1}$ which leads to

$$\frac{1-(\varphi_1+\varphi_2)}{\varphi_2} > \frac{1-(\pi_1+\pi_2)}{\pi_2}$$ and this implies

$$\frac{\varphi_3}{\varphi_2} > \frac{\pi_3}{\pi_2}. \qquad \text{(A26)}$$

Using (A26) it follows that

$$\frac{\varphi_2}{\varphi_1} > \frac{\pi_2}{\pi_1} \qquad \text{(A27)}$$

and therefore this is the only condition to prove.

Now $\lambda > 1$ and $0 < \pi_1 < 1$ and we can write

$$(\lambda - 1)(1-\pi_1) > 0 \qquad \text{(A28)}$$

and

$$(1-\pi_1) + \pi_1 = 1. \qquad \text{(A29)}$$

Adding (A28) and (A29) we get

$$\lambda(1-\pi_1) + \pi_1 > 1. \qquad \text{(A30)}$$

From (A24)

$\frac{\pi_1}{\lambda(1-\pi_1)} = \frac{\varphi_1}{1-\varphi_1}$ and one can derive $\frac{\pi_1}{\lambda(1-\pi_1)+\pi_1} = \varphi_1$.

From (A30) it follows that

$$\frac{\pi_1}{\lambda(1-\pi_1)+\pi_1} = \varphi_1 < \pi_1 < 1. \qquad \text{(A31)}$$

From (A25)

$$\varphi_1 + \varphi_2 < \pi_1 + \pi_2 < 1 \qquad \text{(A32)}$$

$$1 - \pi_1 - \pi_2 < 1 - \varphi_1 - \varphi_2. \qquad \text{(A33)}$$

From (A31) and (A33) we get $\frac{1-\pi_1-\pi_2}{\pi_1} < \frac{1-\varphi_1-\varphi_2}{\varphi_1}$ (A34)

and this implies $-\frac{\pi_2}{\pi_1} < -\frac{\varphi_2}{\varphi_1}$ which gives $\frac{\pi_2}{\pi_1} < \frac{\varphi_2}{\varphi_1}$ or $\frac{\varphi_2}{\pi_2} > \frac{\varphi_1}{\pi_1}$ Q.E.D.

APPENDIX V

PUBLISHED PAPERS

Lall R, Campbell M, Walters S, Morgan K. A review of ordinal regression models applied on health related quality of life assessments. *Statistical Methods in Medical Research* 2000; **11**:49-67
(as enclosed)

Walters S, Campbell M, Lall R. Design and analysis of trials with quality of life as an outcome: A practical guide. *Journal of Biopharmaceutical Statistics* 2001; **11**:155-176

Statistical Methods in Medical Research 2002; **11**: 49–67

A review of ordinal regression models applied on health-related quality of life assessments

R Lall, MJ Campbell Institute of General Practice and Primary Care, University of Sheffield, Community Sciences Centre, Northern General Hospital, Sheffield, UK, **SJ Walters** University of Sheffield, Sheffield Health Economics Group, Sheffield, UK, **K Morgan** Department of Human Sciences, Loughborough University, Loughborough, Leicestershire, UK and **MRC CFAS Co-operative** Institute of Public Health, Cambridge, UK

There has been increasing emphasis in medical research on the design and analysis of quality of life scales. Many quality of life scales are ordinal and statistical methods such as ordinal regression models have been reviewed on a number of occasions. However, when such models are applied, the way the data have been generated is often overlooked. In this paper we illustrate the use of ordinal regression models, in particular the proportional odds model, the partial proportional odds model and the stereotype model in the MRC Cognitive Function and Ageing Study (MRC CFAS). The partial proportional odds and the stereotype models are often under-utilized, perhaps due to the lack of available software. However, in this paper, analysis based on these models has been carried out using the popular statistical software package SAS and macros devised in SAS. Furthermore, bootstrapping techniques have been applied to obtain valid estimates of the standard errors of the parameters in the stereotype model. Strikingly different results were obtained using the different ordinal regression models. We conclude that the way the data have been generated is particularly important for the analysis of quality of life assessments. Different methods of generating scores yield data with different properties. It is now possible to fit a variety of ordinal regression models and so select the appropriate one that correctly models the data.

1 Introduction

There has been an increasing recognition that medical outcomes are not necessarily the most important results in studies that examine the effect of health interventions. This is particularly true for diseases that are presently incurable, such as advanced cancer and chronic diseases of the elderly. It is often the case that two interventions will have very similar medical outcomes, but have different effects on other aspects of people's lives. For this reason there has been increasing emphasis on the use of scales that measure quality of life. It is important that investment in healthcare delivers not only a longer life, but also an improved and maintained quality of life. In conjunction with economic and clinical measures, quality of life outcomes have provided a broader and more accurate assessment of the health status and well-being of patients. In addition, quality of life assessments have provided a means of examining the quality of care given and also have provided utilities such as quality-adjusted life years (QALYs) that aid in policy-making decisions. Quality of life assessments are often measured using questionnaires, and the choice is often between a standard (*generic*) one, which asks about

Address for correspondence: R Lall, Institute of General Practice and Primary Care, University of Sheffield, Community Sciences Centre, Northern General Hospital, Herries Road, Sheffield S5 7AU, UK.

10.1191/0962280202sm271ra

general health and which normally has a history of successful use, and a disease-specific one that has been specifically developed within the therapeutic area in question. The quality of life data can be summarized into a categorical scale that is often either nominal or ordinal in nature. For the analysis of nominal data, standard methods such as the Pearson's chi-squared test, logistic regression models etc, exist which quite adequately provide results and summarize the data. For ordinal scales, more complex statistical methods such as *ordinal regression models*[1] can be employed. When such models are applied to the analysis of quality of life data, the way this data have been generated is often overlooked. Other authors have also highlighted this point. For example, Greenland[6] emphasized that the type of ordinal regression model used for analysis should depend on the way the data have been processed and generated. This is particularly important in the case of quality of life and health status assessments, as different types of data are obtained depending on the biological and sampling processes that generated the data. For instance, in the case of the HADS (Hospital Anxiety and Depression Scale), a response to a question 'I still enjoy things I used to enjoy' can be recorded as: '(0) Definitely as much', '(1) Not as much', '(2) Only a little' or '(3) Hardly at all'. There are in total seven questions and the scores are summed and the final score ranges from 0 to 21. This score is divided into a three-category ordinal scale: 'Normal (<7)'; 'Borderline (8–10)' and 'Clinical depression (or anxiety) (11+)'. The categories on this scale are related to an underlying continuum, which is the score ranging from 0 to 21, and the ordinal variable is termed a '*grouped continuous*' variable.[7,8] The quality of life data obtained in this way are different to that, for instance, in some of the dimensions of the SF-36 quality of life questionnaire.[9] This questionnaire assesses the general health of individuals, and there is a question on health status that asks 'In general would you say your health is 'Excellent', 'Very good', 'Good', 'Fair', 'Poor'?'. Here the rank of the categories is known to exist on a single dimension. Although we can assume the categories are ordered, we do not know the structure of this ordering with respect to a given explanatory variable. Also, although an underlying variable exists when the categories are ordered, it is not directly related to the ordinal categories on the quality of life scale. For this reason Anderson[10] recognized these types of ordered categories as being discrete and referred to the outcome response as a *judged* or an *assessed* variable. Another example of an assessed variable is social class of different occupations, in which the ordering may depend on covariates such as income or level of education.

In general, assessed variables are likely to have greater observer error, and in most cases there is more subjectivity associated with them compared to the grouped continuous variables. It is therefore important to distinguish between the different ordinal quality of life variables, as this has consequences on the choice of ordinal regression models used to analyse the data.

The purpose of this paper is to illustrate the use of ordinal regression models, in particular the *proportional odds model*, the *partial proportional odds models* and the *stereotype model*, in the MRC Cognitive Function and Ageing Study (MRC CFAS).[2] The models are described in Section 2, the data in Section 3 and the fit of the models to the data in Section 4. We conclude with a discussion. Analysis has been carried out using the statistical software package SAS[4] (SAS Institute, Cary, NC; version 6.10 for Windows 95) and macros devised in SAS.[5]

2 Ordinal regression models

Prior to fitting any ordinal regression models, we assessed the general association of the response variable with respect to the covariates using the Cochran–Mantel–Haenszel (row mean scores) statistic as presented by Mantel.[11] This statistic examines the association between the ordinal response variable and one given covariate, while adjusting for the effect of the other covariate by treating it as a stratification variable. The ordinality of the response variable is taken into account by assigning scores to the response categories, forming means, and then examining location shifts of the means across the levels of the rows or *sub-populations* (which result when the levels of the covariates are cross-classified). Furthermore, as it is not clear whether the y-response categories are equally spaced, modified ridit scores were assigned to account for the ordinality. The formulation of the statistic is complex, and the computational details have been omitted as the statistic can be obtained in standard software; in this case PROC FREQ in SAS[12] was used.

2.1 Proportional odds model

The proportional odds model[7] is also known as the *cumulative logit* model. It is the most appropriate method of analysis when one is presented with a grouped continuous response variable. Consider the HADS scale mentioned in the introduction. Let Y denote the response and y_1, y_2 and y_3 indicate the categories of the HADS score: 'Normal (<7)'; 'Borderline (7–10)' and 'Clinical depression (or anxiety) (11+)'. Thus $\Pr(Y = y_j)$ is the probability that a randomly selected individual is in category j. The points '7' and '10' are the *cut-off points*. In generalizing this to a c-point scale, let the response categories be denoted by $y_1, y_2, \ldots, y_c$ and $X_1, X_2, \ldots, X_p$ be a set of explanatory variables or covariates. Taking the proportions $\Pr(Y = y_j) = \pi_j (j = 1, \ldots, c)$ which are based on the marginal totals of a sub-population, one can form cumulative probabilities. Thus for a given sub-population, let $\Pr(Y \leq y_j)$ denote the probability of a response in category y_j or below, ie $\Pr(Y \leq y_j) = \pi_1 + \pi_2 + \cdots + \pi_j$, then $\Pr(Y \leq y_1) \leq \Pr(Y \leq y_2) \cdots \Pr(Y \leq y_c) = 1$ exists, and these cumulative probabilities reflect the ordering in the response categories. The proportional odds model is based on such cumulative probabilities, and this model can be written as:

$$\log\left[\frac{\Pr(Y \leq y_j/X_1 \cdots X_p)}{\Pr(Y > y_j/X_1 \cdots X_p)}\right] = \alpha_j - \left(\beta_1 X_1 + 1\beta_2 X_2 + \cdots + \beta_p X_p\right) \qquad j = 1, 2, \ldots, c-1 \tag{1}$$

Note that the negative sign in the systematic component of (1) makes the sign of each component of β have the usual interpretation in terms of whether the effect is positive or negative. In (1) the regression parameters $\beta_k (k = 1, \ldots, p)$ and α_j are unknown and therefore estimated. The ordinal response categories are monotonically related to an underlying continuous latent variable Z. The relationship between Y and Z is such that the parameters α_j are the division points on the continuum Z and a response in category

y_2, for example, is observed if Z lies between α_1 and α_2. If two adjacent y-response categories are pooled together, or Y is changed by moving a cut-off point, the regression parameters β_k in the model remain unchanged. This property, known as *invariance*, is an attractive feature of the proportional odds model. In practice, however, one uses observed Y values and pooling them will lead to different estimates and inferences of β.

Although the proportional odds model is primarily used in the case where one is presented with a y-response which has an underlying grouped continuous variable, it can also be used in circumstances, for example, when the categories y_j are not directly related to an underlying continuum. In this case, however, the interpretation of the parameters, particularly the $\{\alpha_j s\}$ becomes difficult[7].

The proportional odds model is the most commonly used regression model in the context of analysing ordinal scales,[3] mainly because it provides a single estimate of the log odds ratio over the cut-off points. This estimate is not a weighted average of the cut-off point-specific log odds ratios, but is the optimum estimate obtained using the maximum likelihood or weighted least squares methods. It is ideal in terms of the ease of interpretation of the data and in terms of model parsimony. The $\beta_k(k = 1, \dots, p)$ parameters in model (1) represent the constant cumulative log odds across all the cut-off points for the covariate X_k, having accounted for all the other covariates in the model. The cumulative log odds ratio, λ_k, is obtained by subtracting the log odds (also known as the *logit*) of one row from the log odds of another row.

In SAS, PROC LOGISTIC and PROC CATMOD can be used to fit the proportional odds model. For the analysis presented here, PROC CATMOD was used and the proportional odds model was fitted using the *clogit* link function and an appropriate design matrix.[13] The cumulative logits formed in PROC CATMOD are the reciprocal of those in (1). Note that this procedure only provides the estimates of the $\beta_k(k = 1, \dots, p)$, and the log odds ratios have to be obtained by subtracting the logits of the appropriate rows of interest. The assumption of a constant odds ratio across the cut-off points is assumed for each covariate in model (1), and the $\{\beta_k s\}$ are calculated with this in mind. Prior to fitting a proportional odds model, it is important to carry out a preliminary check of whether the assumption of proportionality is satisfied by each covariate. One way of doing this is to fit a different β_k for each level of the outcome. A different slope model is a starting point for the analysis, and for each covariate the cut-off point-specific β_k parameters together with their standard errors are observed. The homogeneity of the proportional odds ratios over all the cut-off points can be tested using the χ^2-score test statistic.[14] This test is anti-conservative, as it lacks power for moderate departures from the proportional odds assumption, but does highlight major departures. It is also a global test of non-proportionality and does not distinguish heterogeneity associated with individual covariates. In cases where the proportional odds model does not hold for some of the covariates, then alternative models are considered (see below).

2.2 Partial proportional odds models

There is general consensus that the assumptions underlying the proportional odds assumption is quite stringent.[1] This is exacerbated when one considers more than one covariate, and in practice, the chance of all the covariates in the model having proportional odds is likely to be quite rare. The partial proportional odds model permits some covariates to be modelled with the assumption of proportional odds, whilst allowing others to have odds ratios which vary by cut-off point. In general there are two types of partial proportional odds model: the Unconstrained Partial Proportional Odds model and the Constrained Partial Proportional Odds model.[14,15] For the former model the cut-off point-specific odds ratios are obtained for the variables where the odds are thought not to be proportional and a constant odds ratio is obtained for variables where the odds are believed to be proportional. For the latter model, for covariates where the proportional odds assumption does not hold, one may expect a certain 'pattern' in the cut-off point-specific odds ratios, eg a linear trend may be expected in the log odds ratios over the cut-off points. In such a case a set of linear constraints may be placed on the parameter in the model, such that an adequate fit be obtained.

These models are an extension of the proportional odds model. Invariance exists in these models for variables where there are proportional odds and the quality of life scale has an underlying continuum, which is directly related to the y-response categories. Again, due to the way the cumulative probabilities are formed, ordering is inherent in the model irrespective of the relationship of the y-response and the covariate.

2.2.1 *Unconstrained partial proportional odds model*

Let Y be the response variable that has a similar form to that presented in Section 2.1. Then a partial proportional odds model where there are p predictor variables, some of which have proportional odds and some of which have non-proportional odds (say q of them), takes the form:

$$\log\left[\frac{\Pr(Y \leq y_j/X_1 \cdots X_p)}{\Pr(Y > y_j/X_1 \cdots X_p)}\right] = \alpha_j - \Big([\beta_1 X_1 + \gamma_{j1}T_1] + [\beta_2 X_2 + \gamma_{j2}T_2] + [\beta_q X_q + \gamma_{jq}T_q] + \cdots [\beta_p X_p]\Big) \qquad j = 1, 2, \ldots, c-1 \tag{2}$$

Here $X_1, X_2, \ldots X_p$ are the complete set of covariates, and q of these are known to have non-proportional odds, with the remaining having proportional odds. The $\beta_1 \cdots \beta_p$ parameters are the components of each of the covariate-specific log odds, for which proportionality over the cut-off points can be assumed. The $T_1 \cdots T_q$ $(= Y_1 \cdots X_Q)$ exist only for the q variables that have non-proportional odds. Thus $\gamma_{j1} \cdots \gamma_{jq}$ are non-zero for the q-covariates and zero otherwise and are the components of the log odds that vary over the cut-off points.

For model (2), we estimate the $c-1$ intercept parameters, p beta regression parameters that are independent of the cut-off points, and a further $(c-1) \times q$ γ-parameters which are associated with each covariate and cut-off point. For a variable X_m where non-proportional odds exist in relation to the response, $\alpha_j - \beta_m X_m$ is incremented by a regression coefficient $\gamma_{jm}T_m$, which is the effect associated with each jth cumulative logit, having accounted for all the covariates. Note that $\gamma_{1m} = 0$, such that the logit associated with the first cut-off point is based on $\alpha_1 - \beta_m X_m$.

2.2.2 *Constrained partial proportional odds model*

Given the relationship of a covariate and the response is represented with non-proportional odds, then for the individual cut-off point-specific odds ratios, often a certain type of trend may be anticipated, e.g., a linear trend may be expected. In such a case, a set of constraints can be placed on the parameters in the model, so that the trend is taken into account. When the constraints are incorporated into the unconstrained partial proportional odds model, this model becomes:

$$\log\left[\frac{\Pr(Y \leq y_j/X_1 \cdots X_p)}{\Pr(Y > y_j/X_1 \cdots X_p)}\right] = \alpha_j - \Big([\beta_1 X_1 + \beta_2 X_2 + \cdots \beta_p X_p] + \tau_j[\gamma_1 T_1 + \gamma_2 T_2 + \cdots \gamma_q T_q]\Big) \qquad j = 1, 2, \ldots, c-1 \tag{3}$$

The τ_j are fixed scalars that take the form of constraints placed on the parameters. Note that for a given covariate X_m, γ_m does not depend on the cut-off points, but is multiplied by the τ_j for the jth logit. As $\tau_1 = 0$ the first logit is always equal to $\alpha_1 - \beta_m X_m$.

The choice of constraints can be decided upon in several ways; ideally they should be determined using pilot data or one can choose some values which are based on some *a priori* knowledge of the way the odds ratios are likely to behave. However, this is not always possible and some authors[1] have examined the odds ratios obtained from the unconstrained model to determine a set of constraints for the constrained model. This is problematic as one is using the data to estimate the constraints, but may be the only way to obtain the required constants. Regardless of the way the $\{\tau_j s\}$ are obtained, the crucial point is that the same set of constraints has to be assigned to each covariate.

In either model (2) or (3), if we assume that the relationship of the response categories and X_m is represented by non-proportional odds ratios, then the constant component β_m of the log odds ratios and the γ_{jm} (or γ_m) are obtained by fixing and conditioning on all the remaining covariates in the model. $\beta_{jm} = \beta_m + \gamma_{jm}$ refers to the log odds for the jth cut-off point based on the unconstrained model. Similarly $\beta^*_{jm} = \beta^*_m + \tau_j \gamma^*_m$ refers to the log odds based on the constrained partial proportional odds model. The log odds ratios are obtained by subtracting the logits for the appropriate levels (or rows) of the covariate. In the case where X_m is a continuous or ordinal covariate, and there is unit spacing, then for a fixed cut-off point there is a constant log odds ratio when comparing all pairs of adjacent rows, and there are in total $c - 1$ log odds ratios. In the case where X_l is a nominal variable, then, given all the other covariates, the log odds ratios do not only vary by cut-off point, but also vary for the two rows compared, resulting in $(c-1)(r-1)$ log odds ratios. Where a covariate of interest has proportional odds, the $\{\beta_k s\}(k = 1, \ldots, p)$ are interpreted in exactly the same way as those in model (1).

The partial proportional odds models can be fitted in SAS using the PROC CATMOD procedure and the *clogit* link function. The program code differs for each model in the way the design matrix is specified.[13] One can assess whether the proportional odds model is as good a fit as the unconstrained partial proportional odds model by comparing the −2log-likelihoods for the two models. Unfortunately SAS uses weighted least squares estimation for the analysis, and therefore no values for the log-likelihoods are readily available. Thus comparison of the models is made using contrast statements. In the unconstrained model, for a given parameter where different log odds are fitted over the cut-off points the null hypothesis $H_0 : \gamma_{1k} = \gamma_{2k} = \cdots \gamma_{(c-1)k} = 0$ is incorporated into the contrast statement, and this assesses whether the proportional odds model is as good a fit as the unconstrained partial proportional odds model. Likewise, for a covariate where a trend is apparent in the beta parameters and a set of constraints are considered, the test of whether a model using the constraints is as good a fit as a model using the individually estimated log odds can also be obtained using the contrast statements. This test is set up in the unconstrained partial proportional odds model and for a given covariate one assesses the null hypothesis $H_0 : \gamma_{1k} = \tau_1\gamma_k; \gamma_{2k} = \tau_2\gamma_k; \ldots \gamma_{(c-1)k} = \tau_{(c-1)}\gamma_k$.

2.3 Stereotype model

Consider a quality of life scale that assesses 'pain' with respect to some treatment, and assume that the response is recorded on an ordinal scale as 'none', 'mild', 'moderate' or 'severe'. Although the categories are scaled on a single dimension, they are not a discrete version of some continuous variable. A model which assesses the ordinality of the response by looking at the ordering of log odds ratios of the categories is the *stereotype model*.[10] One of the features of this model is that ordering of the response categories with respect to a set of covariates is no longer an assumption but becomes part of a more general model.

The stereotype model is based on the *polytomous regression model*.[1,16] This model is an extension of the logistic regression model and is designed to analyse nominal scales where there are several categories. The logits are formed for this model by comparing each response category with a baseline one, the choice of which is arbitrary and for the analysis presented here, is the first category. Thus the log odds ratio can be represented by a linear model of the form:

$$\log\left[\frac{\Pr(Y = y_j/X_1 \cdots X_p)}{\Pr(Y = y_1/X_1 \cdots X_p)}\right] = \alpha_j + \beta_{j1}X_1 + \beta_{j2}X_2 + \cdots + \beta_{jp}X_p \qquad j = 2, \ldots, c \quad (4)$$

From this model it is clear that the ordinal nature is not accounted for. The ordinality is built into this model by imposing a structure on the log odds $\beta_{jk}(j = 2, \ldots, c;$ $k = 1, \ldots, p)$ such that

$$\beta_{jk} = \phi_j\beta_k \qquad \begin{array}{l} j = 2, \ldots, c \\ k = 1, \ldots, p \end{array} \quad (5)$$

(note: $\phi_1 \equiv 0$ since $\beta_{1k} = 0$).

This results in the stereotype model that takes the form:

$$\log\left[\frac{\Pr(Y = y_j/X_1 \cdots X_p)}{\Pr(Y = y_1/X_1 \cdots X_p)}\right] = \alpha_j + \phi_j\left(\beta_1X_1 + \beta_2X_2 + \cdots + \beta_pX_p\right) \qquad j = 2, \ldots, c \quad (6)$$

Thus, it can be seen that the stereotype model determines a set of weights, the $\{\phi_j s\}$ for the dependent variable and a single parameter β_k for each independent variable. The weights are decided upon for the response variable and are directly related to the effect of the covariates. Thus, when the odds ratios form an increasing trend, the weights can be constrained to be ordered such that

$$0 = \phi_1 \leq \phi_2 \leq \cdots \leq \phi_c \tag{7}$$

Here we can say that the effect of the covariates upon the first odds ratio is less than their effect upon the second and so on, and that, provided constraint (7) holds, model (6) is an ordered regression model. The model fitted does not necessarily require the $\{\phi_j s\}$ to be ordered; whether there is ordering or not is purely determined by the empirical evidence provided by the data.

The weights can be determined in a number of ways. Some authors[6] suggest that they be decided upon *a priori*, either estimated from some pilot data or a suitable set of values be chosen (fixed scores). With such predetermined set of weights, the stereotype model can be fitted in SAS using PROC CATMOD, and the stereotype model remains of a generalised linear form. Hendrickx,[5] however, has designed macros (suitable for use in SAS and in Stata), that fit the stereotype model and estimate the $\{\phi_j s\}$ as a set of parameters in the model (estimated scores). In this case the stereotype model is intrinsically non-linear, but is easily fitted by performing a series of generalized-linear model fits in which the β_k and ϕ_j parameters are alternatively held fixed while the other is estimated. The model produces $(c - 1)$ standard multinomial intercept parameters for the y-response, $(c - 2)$ independent $\{\phi_j s\}$ and a single beta parameter for each independent variable. Although the stereotype model is more flexible than the proportional odds model, it is less parsimonious with the extra weighting parameters. In the case where the weights are estimated, the β_k and ϕ_j parameters are conditional on the estimates, and thus the estimates of the standards errors of these β_k parameters, which assume the $\{\phi_j s\}$ are known, are invalid. Likewise any inference based on the standard errors or the likelihood-based tests is also not correct, and instead we used bootstrap techniques (random sampling with replacement from the original data) to obtain the correct standard errors and tests. Thus using 100 bootstrap samples we refitted the stereotype model to each bootstrap sample and estimated the $\{\beta_k s\}$ and $\{\phi_j s\}$. Using these estimates, six joint distributions for the $\{\beta_k \phi_j\}(j = 1, 2, 3; k = 1, 2)$ were obtained. From each distribution, the mean and standard error were calculated to give the estimate of the log odds ratio and its standard error based on each cut-off point and covariate.

The change in the −2log-likelihood for the polytomous model and the stereotype model provides a way of establishing whether a model with weights is as good a fit as a model where no weights are imposed on a set of covariates. The −2log-likelihoods for these models were also obtained using bootstrap techniques. The null hypothesis was based on the fact that the polytomous model was a good fit to the data. The observed change in the polytomous model and the stereotype model was compared by examining its position in the distribution of the 100 changes in the −2log-likelihoods obtained using the bootstrap samples.

For nominal covariate X_l, β_{2l} is the log odds ratio which is based on comparing the second category with the referent (first) one. To obtain the subsequent cut-off point-specific log odds ratios, $\beta_{jl}(j = 3, \ldots, c)$ the parameter β_l is multiplied by the weights $\phi_j(j = 3, \ldots, c)$ In the case where X_l is a continuous or ordinal and there are r rows with unit spacing, then for a fixed cut-off point j, the log odds ratio is constant when comparing each consecutive pair of rows. Also for a given pair of rows there are a total of $(c - 1)$ log odds ratios. In the case where X_l is nominal, the log odds ratio will vary depending on the two rows compared, as well as over the cut-off points.

3 The data

3.1 The survey

Data (version 4.1) were provided by the Medical Research Council Cognitive Function and Ageing study (MRC CFAS).[2] MRC CFAS commenced as a longitudinal, two-wave (prevalence/incidence), two-stage (screening/assessment) epidemiological study of dementia conducted in six centres throughout England and Wales (urban Liverpool, Newcastle, Nottingham and Oxford, and rural Gwynedd and Cambridgeshire). As the study design was slightly different for the Liverpool centre,[2,17] this centre was omitted in the analysis. At the first visit all respondents were screened with a basic interview covering socio-demographic details, activities of daily living, physical health, measures cognitive function and medication. Subsequently further interviews were carried out (annual follow-up visits and the incidence screen and assessment visits), but these are not detailed here as the analysis only used the first visit data.

At each centre random samples were selected of sufficient size to yield 2500 interviews from individuals aged 65 years and over, with equal numbers in the age groups 65–74 years old and 75 years old and above. The total sample available at baseline was 20 234 for the five centres and there were 17 608 respondents identified as being eligible. Of these 13 006 were interviewed at the initial visit and were regarded as the achieved sample. An outcome that measured the health status of an individual using an ordinal scale was selected for the purpose of the analysis. This outcome was in a form of a question and was asked by an interviewer: 'Would you say that for someone of your age, your own health in general is: 'Excellent', 'Good', 'Fair', 'Poor', 'No answer', 'Not asked' and 'missing'?'. Two categorical covariates 'Have you ever suffered from a heart attack? 'Yes', 'No', 'No answer', 'Not asked' and 'missing'' and 'Do you smoke? 'Yes', 'No', 'No answer', 'Not asked' and 'missing'' were chosen as the independent variables to be used in the models, as these provided a good example of discrimination between the different ordinal regression models outlined.

3.2 Response rates

The number of missing observations for the outcome response was 309 (2.4%), the number of observations where no answer was provided was 61 (0.5%) and the number of respondents who were not asked the health status question was 14 (0.1%). These response categories were ignored as they could not be incorporated into the ordinal scale and could not be analysed using the ordinal regression models. For the 'heart attack' question, the number of subjects who had missing observations was 340 (2.6%),

Table 1 Frequency table for the data from MRC Cognitive and Function Ageing Study (MRC CFAS)

Do you smoke?	Have you had a heart attack?	Rating of health status				
		Excellent	Good	Fair	Poor	Total
Yes	Yes	27 (0.11)	76 (0.31)	101 (0.42)	39 (0.16)	243
Yes	No	402 (0.19)	1050 (0.50)	522 (0.25)	145 (0.07)	2119
No	Yes	83 (0.08)	406 (0.39)	442 (0.42)	114 (0.11)	1045
No	No	1959 (0.21)	4521 (0.50)	2243 (0.25)	405 (0.04)	9128

Parentheses reference the proportions based on the marginal totals.

six (0.05%) respondents did not provide an answer and a further two (0.02%) subjects did not answer the question. For the 'smoke' covariate, there were 392 (3.0%) respondents who had missing data, and four (0.03%) respondents were not asked the question. Missing observations were eliminated from the analysis. Similarly the number of respondents who had 'no answer' or were 'not asked' the question were few and these too were eliminated from the analysis. Thus, although approximately 13 000 elderly people were presented in the 'achieved' sample, complete observations on the response and the covariates were available on 12 535 subjects. The data for the *y*-response (health status) were cross-tabulated with the two covariates of interest to form four sub-populations and are shown in Table 1.

From Table 1, the majority of patients rate their health as 'good' or 'fair' irrespective of whether or not they smoke and whether or not they have had a heart attack. Regardless of whether a respondent smokes or not, there is a tendency for those who have had a heart attack to rate their health lower than those who have not had a heart attack. Regardless of whether a respondent has had a heart attack or not, provided he/she is a smoker there is a greater chance of rating his/her health as 'poor'.

Using the Cochran–Mantel–Haenzel (row mean score) statistic a significant association was found between the general health of the respondent and whether he/she has had a heart attack (after controlling for whether he/she smokes or not—$Q_{SMH} = 190.767$ on 1 d.f.; $P = 0.001$). Likewise there was evidence of a notable association between the health status and whether or not a respondent smokes (after having accounted for the fact that a respondent may or may not have had a heart attack—$Q_{SMH} = 4.212$ on 1 d.f.; $P = 0.04$). Furthermore, there was no evidence that the two covariates of interest were associated ($\chi_1^2 = 0.001$; $P = 0.982$). The main drawback of this method, however, is that it is only capable of displaying the general association. No estimates for the general or cut-off point-specific odds ratios are readily available. Ordinal regression models are a superior way of assessing the relationship between the ordered response and a set of covariates.

4 Fitting the models

4.1 Preliminary analysis and the proportional odds model

Before fitting the proportional odds model, one should check the assumption of proportionality, and so the individual cut-off point-specific cumulative odds ratios were

Table 2 Different slopes models (single cumulative model and three separate logistic regression models)

Variable	Cut-off points					
	(Good, fair, poor) vs excellent		(Fair, poor) vs (excellent, good)		Poor vs (excellent, good, fair)	
	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)
Suffered from a heart attack (yes/no)?	1.0208	0.1024	1.0374	0.0596	0.9656	0.0966
	(1.0459[a]	0.1024)	(1.0348[a]	0.0596)	(0.9664[a]	0.0965)
Do you smoke (yes/no)?	0.1222	0.0590	0.1304	0.0488	0.4592	0.0894
	(0.1246[a]	0.0591)	(0.1242[a]	0.049)	(0.4561[a]	0.0895)

[a]Refers to the analysis carried out by the logistic regression model.

$$\log\left[\frac{\Pr(Y \le y_j/X_1.X_2)}{\Pr(Y > y_j/X_1.X_2)}\right] = \alpha_j - \beta_{j1}(\text{heart attack}) - \beta_{j2}(\text{smoke}); \qquad j = 1, 2, 3.$$

computed. Table 2 illustrates the results where different slopes (based on each cut-off point) were fitted using cumulative probabilities in a single model, and the results of three separate logistic regression models (based on the three cut-off points). It is interesting to note that the results are very similar for the single cumulative model and the three logistic regression models. When comparing the results from these models it can be seen that the standard errors of the log odds estimates are almost identical and there is only a slight variation in the estimates.

By observing the adjusted log odds ratios from the single cumulative model in Table 2, we conclude that there is little difference in the probability of lower ratings of health as opposed to higher in those who may or may not have had a heart attack. For the 'smoke' covariate there appears to be much more variation in the odds ratios. This would suggest that the constant odds assumption is unlikely to be satisfied for the proportional odds model and indeed this was the case as suggested by the χ^2-score test statistics in Table 3 (χ^2-score test statistics: overall, $\chi_4{}^2 = 16.2192$, $P = 0.0027$; 'Heart attack', $\chi_2{}^2 = 0.5804$, $P = 0.7481$; 'smoke', $\chi_2{}^2 = 16.0482$, $P = 0.0003$). Furthermore the proportional odds model was found to be a poor fitting model (chi-squared residual test, $\chi_7{}^2 = 22.76$; P-value $= 0.0019$).

4.2 Unconstrained partial proportional odds model

As the proportional odds assumption does not hold for one of the two covariates in the model, a partial proportional odds model, initially using no constraints, was fitted and the results of this are illustrated in Tables 4 and 5. For the subjects who may have had a heart attack (as opposed to not having had one), a constant adjusted log odds ratio could be assumed across the health status categories. However, the estimated adjusted log odds for those who were/were not smokers varied by cut-off point, and in relation to model (2) are denoted by β_2 (corresponding to the first cut-off point), $\beta_2 + \gamma_{22}$ (corresponding to the second cut-off point) and $\beta_2 + \gamma_{32}$(corresponding to the third cut-off point). Table 4 displays the weighted least-squares estimates and these have been used to obtain the estimates of the log odds ratios together with their standard errors (Table 5). This model accommodates the proportional odds present in

Table 3 Model fitting the proportional odds model

Variable	χ^2-score test statistic	d.f.	*P*-value	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)
Heart attack (yes/no)	0.5804	2	0.7481	1.0241	(0.0553)		
Smoke (yes/no)	16.0482	2	0.0003			0.1460	(0.0426)
Heart attack × Smoke	16.2192	4	0.0027	1.0222	(0.0554)	0.1542	(0.0428)

$$\log\left[\frac{\Pr(Y \leq y_j/X_1.X_2)}{\Pr(Y > y_j/X_1.X_2)}\right] = \alpha_j - \beta_1(\text{heart attack}) - \beta_2(\text{smoke}); \qquad j = 1, 2, 3.$$

Table 4 Unconstrained partial proportional odds model: weighted least squares parameter estimates

Parameters	Estimate ± s.e.	Adjusted heart attack covariate Estimate (s.e.)		Adjusted smoke covariate Estimate (s.e.)		Wald's test statistic	
						$\gamma_1 = \gamma_2 = \gamma_3 = 0$	$\gamma_3 = 40\gamma_2$
α_1	1.8752 ± 0.0379						
α_2	−0.3199 ± 0.0323						
α_3	−2.3404 ± 0.0481						
β		0.5115	(0.0277)	0.0609	(0.0295)		
γ_1				0		$\chi_1^2 = 5.69$	
γ_2				0.00411	(0.0314)	$P = 0.02$	$\chi_1^2 = 0.00$
γ_3				0.1691	(0.0503)		$P = 0.9970$

$$\log\left[\frac{\Pr(Y \leq y_j/X_1.X_2)}{\Pr(Y > y_j/X_1.X_2)}\right] = \alpha_j - \beta_1(\text{heart attack}) - [\beta_2(\text{smoke}) + \gamma_{12}(\text{smoke: health status} > \text{excellent})$$
$$+ \gamma_{22}(\text{smoke: health status} > \text{good}) + \gamma_{32}(\text{smoke: health status} > \text{fair})]; \qquad j = 1, 2, 3.$$

Table 5 Log odds ratios for unconstrained partial proportional odds model

Variable			(Good, fair, poor) vs excellent		(Fair, poor) vs (excellent, good)		Poor vs (excellent, good, fair)	
	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)
	Constant component of log odds ratio across cut-off points		*Increment at cut-off points*					
Suffered from a heart attack (yes/no)?	1.023	0.0554	—	—	—	—	—	—
Do you smoke (yes/no)?	0.1218	0.059	0		0.00822	(0.0628)	0.3382	(0.1006)
			Log odds ratios at cut-off points					
Do you smoke (yes/no)?	—	—	0.1218	(0.059)	0.1300	(0.0991)	0.4600	(0.1281)

the 'heart attack/no heart attack' covariate, and for the non-proportional odds present in the 'smoke/no smoke' covariate. The log odds ratios obtained for this model are very similar to those presented in Table 2. In terms of interpretation, the respondents who have had a heart attack (as opposed to not having had one) are three times as likely to have lower ratings of health as opposed to higher ratings. However, as an underlying continuum directly related to the response categories does not exist for this quality of life scale, the invariance property does not apply. Given a respondent is a smoker (as opposed to not being a smoker), then after adjusting for the fact that he/she may or may not have had a heart attack, the odds of having ('good', 'fair', 'poor') health is 1.1 times that of having ('excellent') health, and this is similar to the odds of having ('fair', 'poor') health versus ('excellent', 'good') health. The adjusted odds of having ('poor') health is 1.6 times that of having ('excellent', 'good', 'fair') health for a smoker (compared to a non-smoker). It is evident that the non-proportionality in the 'smoke' covariate is as a result of the odds ratio obtained at cut-off point 3. The unconstrained partial proportional odds model was found to be a good fitting model to the data (chi-squared test of residuals, $\chi_5{}^2 = 7.14$, P-value $= 0.2102$) and was a better fit than the proportional odds model ($H_0, \gamma_{1k} = \gamma_{2k} = \cdots \gamma_{3k} = 0$; $\chi_2{}^2 = 15.62$; P-value $= 0.01$). Furthermore, it is a more parsimonious model than the model that allowed for separate slopes for the cut-off points for each covariate (since seven parameters are estimated in the unconstrained partial proportional odds model and nine parameters are estimated in the different slopes model).

4.3 Constrained partial proportional odds model

In fitting the constrained partial proportional odds model, the cut-off point-specific log odds ratios are observed from the unconstrained partial proportional odds model. A monotonic trend is apparent in the log odds ratios across the health status categories in relation to the covariate that assesses whether respondents smoke or not smoke. In order to simplify the interpretation, a constraint can be placed on the parameters (leading to the formation of the constrained partial proportional odds model). Thus whilst a proportional odds is apparent in the variable assessing 'heart attack', the 'smoke' variable has odds ratios which follow an increasing trend over the cut-off points. Based on these, the following constraints were chosen: $\tau_{12} = 0$; $\tau_{22} = 1$; $\tau_{32} = 40$. These formed the following log odds: $\beta_2 + 0 \cdot \gamma_2 T_2$, $\beta_2 + 1 \cdot \gamma_2 T_2$ and $\beta_2 + 40 \cdot \gamma_2 T_2$. The parameter estimates and the log odds ratios are presented in Tables 6 and 7. The interpretation of the parameter estimates for this model is very similar to that for the unconstrained partial proportional odds model. The constrained partial proportional odds model was found to be a good fit model (test of residuals: $\chi_6{}^2 = 7.16$; P-value $= 0.3036$). This model was found not to be significantly different from the unconstrained model ($H_0, \gamma_3 = \tau_3\gamma_2$; $\chi_1{}^2 = 0.00$; P-value $= 0.970$) and therefore in terms of model parsimony was the preferred model (as only six parameters were estimated, as the constraints are not considered model parameters in this case).

4.4 Polytomous model

Before fitting the stereotype model, we examined the log odds ratios (and their standard errors) from the polytomous model. In both models the referent category was chosen to be 'excellent', and therefore the cut-off point-specific odds ratios are based on

Table 6 Constrained partial proportional odds model: weighted least squares parameter estimates

Parameters	Estimates (s.e.)		Adjusted heart attack covariate		Adjusted Smoke covariate	
			Estimates (s.e.)		Estimates (s.e.)	
α_1	1.8751	(0.0357)				
α_2	−0.3199	(0.0315)				
α_3	−2.3404	(0.0479)				
β			0.5115	(0.0277)	0.0608	(0.0217)
γ (constraint parameter)					0.00423	(0.00107)

$$\log\left[\frac{\Pr(Y \leq y_j/X_1.X_2)}{\Pr(Y > y_j/X_1.X_2)}\right] = \alpha_j - \beta_1(\text{heart attack}) - [\beta_2(\text{smoke}) + \tau_1\gamma_2(\text{smoke: health status>excellent})$$

$$+ \tau_2\gamma_2(\text{smoke: health status>good}) + \tau_3\gamma_2(\text{smoke: health status>fair})]; \quad j = 1, 2, 3.$$

Constraints: $\tau_1 = 0$; $\tau_2 = 1$; $\tau_3 = 40$.

Table 7 Log odds ratios for constrained partial proportional odds model

Variable			(Good, fair, poor) vs excellent		(Fair, poor) vs (excellent, good)		Poor vs (excellent, good, fair)	
	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)
	Constant component of log odds ratio across cut-off points				*Increment at cut-off points*			
Suffered from a heart attack (yes/no)?	1.023	(0.0554)	—		—	—	—	—
Do you smoke (yes/no)?	0.1216	(0.0434)	0		0.00846	(0.00214)	0.3384	(0.0856)
					Log odds ratios at cut-off points			
Do you smoke (yes/no)?	—	—	0.1216	(0.0434)	0.1300	(0.0435)	0.4600	(0.0455)

'good' versus 'excellent' (cut-off point 1), 'fair' versus 'excellent' (cut-off point 2) and 'poor' versus 'excellent' (cut-off point 3). The weighted least squares estimates and the adjusted odds ratios for both covariates using the polytomous model are displayed in Tables 8 and 9, respectively. Using the test statistic values, we found that the cut-off point-specific adjusted log odds ratios for the 'heart attack' variable are significantly different from one another (Wald's test statistics, $\beta_1 = \beta_2$, $\chi_1{}^2 = 150.78$, $P < 0.0001$; $\beta_2 = \beta_3$, $\chi_1{}^2 = 11.68$, $P = 0.0006$). A constant adjusted log odds ratio can be assumed for the 'good' versus 'excellent' and 'fair' versus 'excellent' cut-off points for the smokers versus non-smokers (Wald's test statistics, $\beta_1 = \beta_2$, $\chi_1{}^2 = 0.12$, $P = 0.73$). However, a different odds ratio has to be assumed for the 'poor' versus 'excellent' categories (Wald's test statistics, $\beta_2 = \beta_3, \chi_1{}^2 = 19.43$, $P < 0.0001$). Note that this is in contradiction to the conclusions drawn from the proportional odds model and the partial proportional odds models. Thus using the polytomous model, for the respondent who has suffered from a heart attack, the adjusted odds of having 'good' health is

Table 8 Polytomous logistic regression model: weighted least squares parameter estimates

Parameters	Estimate	(s.e.)	Adjusted heart attack covariate		Adjusted smoke covariate	
			Estimate	(s.e.)	Estimates	(s.e.)
α_1	1.1957	(0.0580)				
α_2	0.9110	(0.0587)				
α_3	−0.4108	(0.0724)				
β_1			0.3098	(0.0542)	0.0429	(0.0313)
β_2			0.7197	(0.0541)	0.0524	(0.0329)
β_3			0.8955	(0.0669)	0.2656	(0.0509)

$$\log\left[\frac{\Pr(Y = y_j/X_1.X_2)}{\Pr(Y = y_1/X_1.X_2)}\right] = \alpha_j + \beta_{j1}(\text{heart attack}) + \beta_{j2}(\text{smoke}); \quad j = 2, 3, 4.$$

Table 9 Polytomous model

Variable	Cut-off points						Wald's test statistic	
	Good vs excellent		Fair vs excellent		Poor vs excellent			
	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)	$\beta_1 = \beta_2$	$\beta_2 = \beta_3$
Suffered from a heart attack (yes/no)?	0.6196	(0.1084)	1.4394	(0.1082)	1.791	(0.1338)	$\chi_1^2 = 32.61$ $P = 0.0000$	$\chi_1^2 = 150.78$ $P = 0.0000$
Do you smoke (yes/no)?	0.0858	(0.0626)	0.1048	(0.0698)	0.5312	(0.1018)	$\chi_1^2 = 1.88$ $P = 0.1706$	$\chi_1^2 = 0.12$ $P = 0.7323$

approximately twice that of having 'excellent' health. The adjusted odds of having 'fair' health is approximately four times that of having 'excellent' health and the adjusted odds of having 'poor' health is approximately six times of having 'excellent' health. For the smokers the adjusted odds of having 'good' or 'fair' health is approximately 1.1 times that of having 'excellent' health, and the adjusted odds of having 'poor' health is approximately 1.7 times of having 'excellent' health. The polytomous model was found to be a good fit model (test of residuals, $\chi_3^2 = 6.37$, P-value = 0.0949), although it lacks parsimony.

The difference in the results produced by the proportional odds models and the polytomous model can be explained by assessing the proportions in Table 1. It is evident that the smoker *vs* non-smoker contrast within each of the 'heart attack' subpopulations is very similar across the health status categories. Thus for the 'smoke' covariate, the difference in the two types of models is due to the way the logits are formed. In fact, the log odds ratios are quite similar for this covariate given the two cumulative models and polytomous model. When assessing the proportion of those who have had a heart attack *vs* not having had one, within each of the levels of the smoke covariate, there is a notable difference. The ratio of those who have had a heart attack compared to those who have not had a heart attack increases over the health status categories for both levels of the 'smoke' covariate and this is manifested in the polytomous model. The effect of accumulating the probabilities over the cut-off points removes the increase in the odds ratios to reveal proportional odds.

4.5 Stereotype model

As stated earlier, the polytomous model can be simplified to become the stereotype model. The weights were estimated as parameters using maximum likelihood estimation in model (6), together with the α_j and β_k parameters are presented in Table 10. Differences between the ϕ_j scale values indicates how the log odds of one health status category versus another is affected by having/not having a heart attack and whether respondents smoke or not. The impact of the independent variables on a log odds between adjacent categories is largest for 'good' versus 'poor' health status where the difference between the scale values is 0.439. The smallest impact is on the 'fair' versus 'poor' health status categories, with a difference of only 0.223. The impact of having a heart attack on the logit of having 'excellent' versus 'good' health status is 0.6147 ($\phi_1\beta_1$), resulting in the log odds ratio. Using the remaining weights and the β_k values, the cut-off point-specific log odds ratios can be obtained in a similar way for both covariates. The odds ratios were found not to be much different to those obtained using the polytomous model (see Table 11). The interpretation of these log odds is also similar to that of the polytomous model. As the weights are ordered in a monotone fashion we can assume that there is an ordering in the y-response with respect to the covariates. The observed −2log-likelihood for the nine-parameter polytomous model was 29329.16 compared to a −2log-likelihood of 29 343.73 for the seven-parameter stereotype model. The observed change in the −2log-likelihood values of these two models was 14.57. Comparing this with its position in the bootstrap distribution of 100 changes in −2log-likelihoods of the two models, it was evident that the constrained model was not as good a fit as the polytomous model. Assessing the distribution of the change in the −2log-likelihood values, it was found that 45.5% of the bootstrap sample values lay below the observed value and 54.6% lay above. This implied that the P-value was approximately 0.5, indicating that the null hypothesis (which was based on the fact that the polytomous model was a good fit to the data) could not be rejected. Thus the stereotype model was found to be a poor fit compared to the polytomous model.

Table 10 Parameter estimates using the bootstrap techniques: stereotype model

Parameters	Adjusted heart attack covariate		Adjusted smoke covariate	
	Estimate	(s.e.)	Estimate	(s.e.)
β	1.8468	(0.1346)	0.2703	(0.0805)
ϕ_1	0		0	
ϕ_2	0.3407	(0.0426)	0.3407	(0.0426)
ϕ_3	0.7770	(0.0534)	0.7770	(0.534)
ϕ_4	1		1	

$$\log\left[\frac{\Pr(Y = y_j/X_1 \cdot X_2)}{\Pr(Y = y_1/X_1 \cdot X_2)}\right] = \alpha_j + \phi_j(\beta_1(\text{heart attack}) + \beta_2(\text{smoke})); \quad j = 2, 3, 4.$$

Table 11 Stereotype model

Variable	Cut-off points					
	Good vs excellent		Fair vs excellent		Poor vs excellent	
	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)	ln(O.R.)	s.e. ln(O.R.)
Suffered from a heart attack (yes/no)?	0.6305	(0.0980)	1.4328	(0.1189)	1.8468	(0.1346)
Do you smoke (yes/no)?	0.0920	(0.0295)	0.2082	(0.0575)	0.2703	(0.0805)

5 Discussion

An alternative ordinal regression model, which was not considered in the analysis is the *Fienberg's continuation ratio model.*[18] This model is usually relevant when an ordinal quality of life scale may be thought of as a progression through various stages, so that people start with 'excellent' and deteriorate to 'poor' and are unlikely to reverse this trend. Such data usually resemble failure-time data or outcomes measuring threshold points. The data in this paper were not of this type and therefore the continuation ratio model was considered as being irrelevant.

Irrespective of the modelling techniques, the response variable in a quality of life scale can essentially arise in one of two ways: (a) where there are clearly ordered categories for which there is a single underlying latent variable; or (b) where the categories are discrete and for which ordering may depend on covariate information.

Having established how the data are generated, then one is in a position to decide which model will be most appropriate in terms of analysing the data. Given the y-response is a grouped continuous response variable, the proportional odds and partial proportional odds models are often the most applicable due to the assumptions these models make. In the case of the data presented, the proportional odds model was found to be a poor fitting model and one could have used other forms of strategies. These would include using a different link function in the model, eg the log–log function would produce a response function that was non-symmetric, or alternatively one could include additional terms in the model, such as the interaction term (although in this case the interaction was found to be non-significant). Generally, however, as the covariates increase in a proportional odds model, the lack of fit increases but is compensated by the parsimony of the model. The unconstrained partial proportional odds model is a better fitting model than the proportional odds model, although the parameters increase at a drastic rate as the number of covariates and number of y-response categories increase. The constrained partial proportional odds model would probably be the most ideal model, given a set of covariates and a k-ordered group continuous response variable. However, obtaining the constraints is somewhat problematic, especially if there are a large number of covariates with non-proportional odds. Presently there is no method available for estimating the constraints, and one can only use fixed constraints that have been determined prior to fitting the model.

When the ordered categories in the response variable are of a discrete nature, and there is no directly related underlying continuum, the interpretation of the parameters in the proportional odds and partial proportional odds models becomes difficult. Ideally one would fit the polytomous model when presented with such response data. However, although this produces a good fitting model, it is at the cost of estimating a large number of parameters. As the number of covariates increases in the model, the stereotype model becomes more parsimonious and the estimation of the weights is not problematic, regardless of the number of covariates presented.

The stereotype model would ideally be the most favourable of all these models given that one is presented with an assessed response. However, results from our data demonstrate that the stereotype model was not as good a fit as the polytomous model and the fit of the stereotype model was further illustrated using the Akaike Information Criterion[4,19] ($\text{AIC} = -2 \times \text{log-likelihood} + 2 \times p$, where p is the number of parameters in the model). This criterion adjusts the −2log-likelihood statistic for the number of terms in the model and the number of observations used. It is clear that there is not much difference in fit for the proportional odds or stereotype model (AIC for proportional odds model = 29353.837; AIC for the stereotype model = 29353.73 and the AIC for the polytomous model = 29339.163).

Given the results above, despite the number of parameters estimated, the polytomous model is taken to be the most appropriate model that summarized the data. It is found to fit the data well and, more importantly, it allows for the processes that generate the data. Thus using this model we can conclude that smoking has a greater impact on health status than does having had a heart attack. The odds of having a lower rating of health increase dramatically if one smokes (adjusting for whether they have had a heart attack or not) compared to a non-smoker.

It is now computationally possible to fit most, if not all, the different ordinal regression models using routine statistical packages. There is therefore no reason why one should not account for as much information as possible regarding the data. In this paper, we have attempted to illustrate that the way the data have been generated can be accommodated in a given ordinal regression model and this provides more accurate and refined results.

Acknowledgement

MRC Cognitive Function and Aging study is supported by major awards from the Medical Research Council and the Department of Health. The MRC CFAS website can be obtained on: www.mrc-bsu.cam.ac.uk/cfas/

References

1 Ananth C, Kleinbaum D. Regression models for ordinal responses: a review of methods and applications. *International Journal of Epidemiology* 1997; 26: 1323–33.

2 MRC CFAS Study. The MRC Multicentre Study of Cognitive Function and Ageing: a EURODEM incidence study in progress. *Neuroepidemiology* 1992; **11**: 37–43.

3 Scott S, Goldberg M, Mayo N. Statistical assessment of ordinal outcomes in comparative studies. *Journal of Clinical Epidemiology* 1997; 50: 45–55.

4 *SAS/STAT User's Guide, Version 6*, 4th edn, Vol 2. Cary, NC: SAS Institute.
5 Hendrickx J. Special restrictions in multinomial logistic regression. *Stata Technical Bulletin* 2000; STB-56: 18–26.
6 Greenland S. Alternative models for ordinal logistic regression. *Statistics in Medicine* 1994; 13: 1665–77.
7 McCullagh P. Regression models for ordinal data (with discussion). *Journal of Royal Statistical Society Series B* 1980; 42: 109–42.
8 Engel J. Polytomous logistic regression. *Stat Neerlandica* 1988; 42: 233–52.
9 Ware J, Snow K, Kosinski M, Gandek B. *SF-36 health survey manual and interpretation guide*. Boston, MA: New England Medical Centre, the Health Institute, 1993.
10 Anderson JA. Regression and ordered categorical variables (with discussion). *Journal of Royal Statistical Society Series B* 1984; 46: 1–30.
11 Mantel N. Chi-squared tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association* 1963; 58: 690–700.
12 *SAS Procedures Guide, Version 6*, 4th edn, Vol 2. Cary, NC: SAS Institute.
13 Stokes M, Davis C, Koch G. *Categorical data analysis using the SAS system*. Cary, NC: SAS Institute, 1995.
14 Peterson B, Harrell F. Partial proportional odds model for ordinal response variables. *Applied Statistics* 1990; 39: 205–17.
15 Peterson B, Harrell F. Partial proportional odds models and the LOGIST procedure. In *SAS Users Group International, Thirteenth Annual Conference*, Orlando FL. Cary, NC: *SAS* Institute, 1988, pp 952–56.
16 Agresti A. Tutorial on modelling ordered categorical response data. *Psychological Bulletin*, 1989; 105: 290–301.
17 MRC CFAS. Cognitive function and dementia in six areas of England and Wales: the distribution of MMSE and prevalence of GMS organicity level in the MRC CFA Study. *Psychological Medicine* 1998; 28: 319–35.
18 Everitt BS, Dunn G. *Advanced Methods of Data Exploration and Modelling*. London: Heinemann Educational, 1983: 164–75.
19 *STATA manual, Version 7*. College Station, TX: Statacorp, 2001.