A Logistic Regression Model of the Decision of Volunteers to Enter a Sports Coach Education Programme

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the School of Educational Studies, University of Surrey

© Brian E. Wilson 2002
ACKNOWLEDGEMENTS

The author wishes to acknowledge the immense encouragement and suggestions that he received throughout the period of research from the academic and administrative staff of the University of Surrey, in particular Professor Karen Evens, Professor Peter Jarvis, Dr Tom Black and Dr David Gray of the Department of Educational Studies. In particular, the author will always remember the numerous statistical discussions held with Dr Tom Black, including the many amusing anecdotes that always flowed. The work on the measurement of readability would not have been possible without the initial suggestions of Professor Colin Harrison of the School of Education at the University of Nottingham.

Grateful thanks are made to the Chairmen and Committees of the two athletic clubs for releasing their membership lists and to the committee of the Surrey County Athletic Association for releasing the list of coaches. Without these lists, the research would not have been possible. It is hoped that some of the findings of this research may be of use to the these clubs and to the sport of athletics.

But, above all, the author owes an immense debt of gratitude to the individual coaches, athletes, volunteers and parents connected with athletics who so readily agreed to be interviewed and spoke so frankly, sometimes on very personal subjects. It was astonishing how hospitable people were and how much time they were prepared to give, often by finding precious time in a very busy life. Their collective inputs can never be adequately repaid.

The author must thank the many dedicated athletes and coaches that he has had the privilege of working with for so many years, together with the many colleagues that he had in a variety of other voluntary organisations. It was the knowledge and interests that they created in him that instituted the study and made the completion of the research possible. May they all find as much happiness and stimulation in the voluntary sector as he has done over 50 years.

The research could not have been accomplished without the continuous support and encouragement of Joan Wilson, his wife, and their children, Caroline and Andrew. Only a stable and loving family could have shared and absorbed the intensities and the pressures encountered during the five years of the project.
SUMMARY

200,000 voluntary coaches ensure that the large investments in sports facilities in the UK are effectively used and that the athletes experience a healthy life. In a socio-economic period in which volunteering may be diminishing, more coaches to successfully implement the national strategy for sport. Yet very little is known about how individuals, in particular parents, decide to become voluntary coaches. The research set out to test if a quantitative model could represent such a decision process. It formulates a model of the process and tests the model against field data.

From Social Exchange Theory, a conceptual model was developed of the decision process to enter a coach education programme. It was clear that the process was multi-dimensional. The research then showed how this multi-dimensional conceptual model could be mathematically represented by a logistic regression model. Published work on volunteering, sport and coaching was reviewed. From this literature review, a set of potential explanatory variables was obtained which previous researchers had suggested to be involved in decision processes involving volunteering and coaching. Hypotheses were made for these potential explanatory variables.

Data was collected by personal interview of a sample of 112 individuals associated with athletics in the UK. The sample was taken from qualified, voluntary athletic coaches, from adult members of two amateur athletic clubs, and from parents of children in these clubs. The building of the logistic regression model showed the relative importance of each of the explanatory variables to the decision process. It also showed whether or not variables were independent or associated with another variable. Tests were conducted to indicate how well the mathematical model fitted the data and, hence, how well the conceptual model represented the sample. The use of logistic regression allowed the potential explanatory variables to be ranked in order of influence on the decision to become a coach. It also allowed the hypotheses to be tested.

Qualitative information was also obtained from the interviews. These allowed the logistic regression model to be triangulated. The qualitative data gave further insights into the perceptions of adults, especially parents, and of coaches. It also indicated the key events which triggered a decision to coach.

The logistic regression model provided a satisfactory fit to the data indicating the adequacy of the conceptual model. It indicates that the two genders have differing sets of significant explanatory variables. Some variables were found to be more complex than suggested by the published literature, others to be non linear and some not significant. The degree of fit suggested that one or more unknown, but significant, variables still need to be identified.

The research suggests how the models could be improved and developed. There are implications for the future recruitment, training and retention of voluntary athletic coaches. By extension, the conceptual and mathematical models can be adapted for other adult education decision processes where the choice of course is voluntary.
CONTENTS

ACKNOWLEDGEMENTS

SUMMARY

PAGE

1. INTRODUCTION
   1.1. The Research Question 1
   1.2. Background 6
   1.3. Research Basis 10

2. VOLUNTEERING
   2.1. Introduction 16
   2.2. Definition 16
   2.3. Volunteering and Society 18
   2.4. Altruism 22
   2.5. Affiliation 25
   2.6. Expectancy Theory 28

3. SPORTS COACHING
   3.1. Introduction 32
   3.2. Definitions 34
   3.3. Sport and Volunteering 36
   3.4. Sport, Government and Coaching 43

4. THE UK ATHLETICS COACH EDUCATION PROGRAMME
   4.1. Introduction 47
   4.2. Recruitment 48
   4.3. The Athletics Coach Education Programme 49

5. DECISION MODELS
   5.1. Social Exchange Theory 53
   5.2. The Decision Threshold 59
   5.3. The Sports Model 64

6. STATISTICAL METHODOLOGY
   6.1. Introduction 70
   6.2. Summary of Chapter 6 & 7 72
   6.3. Generalised Linear Models (GLM) 78
   6.4. GLM Components 81
   6.5. The Response and Explanatory Variables 82
   6.6. Potential Forms of GLM 84
7. LOGISTIC REGRESSION
   7.1. The Logit Function 87
   7.2. Assumptions 91
   7.3. Goodness of Fit 95
   7.4. Dispersion 104
   7.5. Collinearity 108
   7.6. Dummy Variables 113

8. DATA COLLECTION
   8.1. National Population 115
   8.2. Research Population 119
   8.3. Sample Size 121
   8.4. Sampling 121
   8.5. Data Weights & Missing Data 124
   8.6. Approach to Respondents 125
   8.7. Conduct of Interviews 127
   8.8. Response Rates and Reasons 130
   8.9. Sample Validity 135

9. INTERVIEWS & THE QUESTIONNAIRE
   9.1. Introduction 137
   9.2. Quantitative Variables, Instruments & Hypotheses 141
   9.3. Qualitative Confirmatory Information 166
   9.4. Background Information 170

10. THE DATA & MODELLING
    10.1 Missing Data 171
    10.2. Data Summaries 171
    10.3. Data Vetting 174
    10.4. Initial Collinearity Test 194
    10.5. Model Building 196
    10.6. Confirmation & Diagnostics 207
    10.7. Relative Significance of Variables 215
    10.8. Coefficients and Standard Errors 221
    10.9. Comparisons with Hypotheses 227
    10.10. Coefficient Signs 232
    10.11. Variables Not in the Model 234
    10.12. Conclusions from the Thesis Model 243

11. GENDER MODELS
    11.1 Introduction 245
    11.2. The Male Model 246
    11.3. The Female Model 253
    11.4. Gender Model Comparisons 260
    11.5. Conclusions from Gender Models 263
### 12. QUALITATIVE EVIDENCE FROM INTERVIEWS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.1. Introduction</td>
<td>266</td>
</tr>
<tr>
<td>12.2. Evidence from Non-Coaches</td>
<td>267</td>
</tr>
<tr>
<td>12.3. Evidence from Coaches</td>
<td>273</td>
</tr>
<tr>
<td>12.4. ‘Key’ Events</td>
<td>278</td>
</tr>
<tr>
<td>12.5. Comparing the Qualitative Evidence</td>
<td></td>
</tr>
<tr>
<td>with the Quantitative Models</td>
<td>280</td>
</tr>
</tbody>
</table>

### 13. RESEARCH OUTCOMES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.1. Answering the Research Question</td>
<td>284</td>
</tr>
<tr>
<td>13.2. Limitations of the Research</td>
<td>286</td>
</tr>
<tr>
<td>13.3. Potential Improvements to the Research</td>
<td>288</td>
</tr>
<tr>
<td>13.4. Implications for Sport in the UK</td>
<td>292</td>
</tr>
<tr>
<td>13.6. Implications for Adult Education in the UK</td>
<td>298</td>
</tr>
</tbody>
</table>

### REFERENCES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Initial Contact Letter</td>
<td>305</td>
</tr>
<tr>
<td>B. Reply slip</td>
<td>306</td>
</tr>
<tr>
<td>C. Interview Questionnaire &amp; Show Cards</td>
<td>307</td>
</tr>
<tr>
<td>D. Position Paper of the Volunteerism Commission</td>
<td>326</td>
</tr>
<tr>
<td>NO.</td>
<td>DIAGRAMS</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
</tr>
<tr>
<td>1.3.1</td>
<td>KEY STAGES &amp; PROCESS OF RESEARCH</td>
</tr>
<tr>
<td>5.1.1</td>
<td>SOCIAL EXCHANGE THEORY</td>
</tr>
<tr>
<td>5.1.2</td>
<td>SOCIAL EXCHANGE THEORY IN VOLUNTARY SPORT</td>
</tr>
<tr>
<td>5.1.3</td>
<td>SOCIAL EXCHANGE THEORY FOR COACHING</td>
</tr>
<tr>
<td>5.2.1</td>
<td>CONCEPTUAL MODEL OF THE RECREATIONAL ACTIVITY ADOPTION</td>
</tr>
<tr>
<td>5.2.2</td>
<td>MODIFIED BRANDENBURG MODEL - THE SPORT MODEL</td>
</tr>
<tr>
<td>5.3.1</td>
<td>SPORTS MODEL ADAPTED FOR COACHING</td>
</tr>
<tr>
<td>7.1.1</td>
<td>THE LOGISTIC FUNCTION</td>
</tr>
<tr>
<td>7.1.2</td>
<td>THE ODDS FUNCTION</td>
</tr>
<tr>
<td>7.1.3</td>
<td>THE LOG ODDS OR LOGIT FUNCTION</td>
</tr>
<tr>
<td>7.1.4</td>
<td>HOMOSCEDASTICITY</td>
</tr>
<tr>
<td>7.1.5</td>
<td>HETEROSCEDASTICITY</td>
</tr>
<tr>
<td>7.3.1</td>
<td>DATA FOR TABLE 7.3.1</td>
</tr>
<tr>
<td>7.3.2</td>
<td>DATA FOR TABLE 7.3.2</td>
</tr>
<tr>
<td>7.3.3</td>
<td>SIMULTANEOUS INTRODUCTION</td>
</tr>
<tr>
<td>7.3.4</td>
<td>SEQUENTIAL INTRODUCTION</td>
</tr>
<tr>
<td>7.3.5</td>
<td>DUMMY VARIABLE CODING</td>
</tr>
<tr>
<td>10.3.1</td>
<td>FLESCH SCORES BY PUBLICATION &amp; PERSPECTIVE</td>
</tr>
<tr>
<td>10.3.2</td>
<td>DIAGRAM 10.3.1. REDRAWN</td>
</tr>
<tr>
<td>10.3.3</td>
<td>LINEARITY CHECK - CARS</td>
</tr>
<tr>
<td>10.3.4</td>
<td>LINEARITY CHECK - CARS V. PROBABILITY OF OTHER SPORTS</td>
</tr>
<tr>
<td>10.3.5</td>
<td>LINEARITY CHECK - TIME</td>
</tr>
<tr>
<td>10.3.6</td>
<td>LINEARITY CHECK - DISTANCE</td>
</tr>
<tr>
<td>10.3.7</td>
<td>LINEARITY CHECK - HOME</td>
</tr>
<tr>
<td>10.3.8</td>
<td>LINEARITY CHECK - HOME (Categorical)</td>
</tr>
<tr>
<td>10.3.9</td>
<td>LINEARITY CHECK - BOYS NOT IN ATHLETICS</td>
</tr>
<tr>
<td>10.3.10</td>
<td>LINEARITY CHECK - BOYS IN ATHLETICS</td>
</tr>
<tr>
<td>10.3.11</td>
<td>LINEARITY CHECK - GIRLS NOT IN ATHLETICS</td>
</tr>
<tr>
<td>10.3.12</td>
<td>LINEARITY CHECK - GIRLS IN ATHLETICS</td>
</tr>
<tr>
<td>10.3.13</td>
<td>LINEARITY CHECK - MEN NOT IN ATHLETICS</td>
</tr>
<tr>
<td>10.3.14</td>
<td>LINEARITY CHECK - MEN IN ATHLETICS</td>
</tr>
<tr>
<td>10.3.15</td>
<td>LINEARITY CHECK - WOMEN NOT IN ATHLETICS</td>
</tr>
<tr>
<td>10.3.16</td>
<td>LINEARITY CHECK - WOMEN IN ATHLETICS</td>
</tr>
<tr>
<td>10.3.17</td>
<td>LINEARITY CHECK - PROPORTION WORK YEARS COMPETING</td>
</tr>
<tr>
<td>10.3.18</td>
<td>LINEARITY CHECK - PROPORTION WORK YEARS COMPETING (Categorical)</td>
</tr>
<tr>
<td>10.3.19</td>
<td>LINEARITY CHECK - PARTICIPATION RATE</td>
</tr>
<tr>
<td>10.3.20</td>
<td>LINEARITY CHECK - PROPORTION SPORT IN WORK</td>
</tr>
<tr>
<td>10.3.21</td>
<td>LINEARITY CHECK - % SPORT IN WORK</td>
</tr>
<tr>
<td>10.3.22</td>
<td>LINEARITY CHECK - YEARS AT SCHOOL</td>
</tr>
<tr>
<td>10.3.23</td>
<td>LINEARITY CHECK - YEARS AT SCHOOL (Categorical)</td>
</tr>
<tr>
<td>10.3.24</td>
<td>LINEARITY CHECK - YEARS FE/HE</td>
</tr>
<tr>
<td>10.3.25</td>
<td>LINEARITY CHECK - YEARS FE/HE (Categorical)</td>
</tr>
<tr>
<td>10.3.26</td>
<td>LINEARITY CHECK - FLESCH EASE OF READING SCORE</td>
</tr>
<tr>
<td>10.3.27</td>
<td>LINEARITY CHECK - WORD COMPLEXITY</td>
</tr>
<tr>
<td>10.3.28</td>
<td>LINEARITY CHECK - SENTENCE LENGTH</td>
</tr>
<tr>
<td>10.3.29</td>
<td>LINEARITY CHECK - ATTITUDE TO VOLUNTEERING</td>
</tr>
<tr>
<td>10.3.30</td>
<td>LINEARITY CHECK - ATTITUDE TO VOLUNTEERING (Categorical)</td>
</tr>
<tr>
<td>10.3.31</td>
<td>LINEARITY CHECK - ATTITUDE TO PAYING COACHES</td>
</tr>
<tr>
<td>10.3.32</td>
<td>CLASSIFICATION, BY ACTUAL AND PREDICTED &quot;Coach&quot;</td>
</tr>
<tr>
<td>10.3.33</td>
<td>DISTRIBUTION OF DEVIANCES RELATED TO NORMAL DISTRIBUTION(Thesis Model)</td>
</tr>
<tr>
<td>10.3.34</td>
<td>STANDARDISED COEFFICIENTS - THESIS MODEL - BY RANK</td>
</tr>
<tr>
<td>10.3.35</td>
<td>-2LL CONTRIBUTION TO FINAL MODEL - BY STANDARDISED COEFFICIENT RANK</td>
</tr>
<tr>
<td>10.3.36</td>
<td>REDUCTION IN % OF &quot;Coach&quot; CORRECTLY ASSIGNED BY VARIABLE</td>
</tr>
<tr>
<td>10.3.37</td>
<td>LOGITS OF VARIABLES RELATED TO DEPENDANTS INCLUDED IN THE THESIS MODEL</td>
</tr>
<tr>
<td>10.3.38</td>
<td>CLASSIFICATION, BY ACTUAL AND PREDICTED &quot;Coach&quot; (Male Model)</td>
</tr>
<tr>
<td>10.3.39</td>
<td>DISTRIBUTION OF DEVIANCES RELATED TO NORMAL DISTRIBUTION (Male Model)</td>
</tr>
<tr>
<td>10.3.40</td>
<td>STANDARDISED COEFFICIENTS - MALE MODEL</td>
</tr>
<tr>
<td>10.3.41</td>
<td>-2LL CONTRIBUTION TO MALE MODEL</td>
</tr>
<tr>
<td>10.3.42</td>
<td>CLASSIFICATION, BY ACTUAL AND PREDICTED &quot;Coach&quot; - FEMALE MODEL</td>
</tr>
<tr>
<td>10.3.43</td>
<td>DISTRIBUTION OF DEVIANCES RELATED TO NORMAL DISTRIBUTION - FEMALE MODEL</td>
</tr>
<tr>
<td>10.3.44</td>
<td>STANDARDISED COEFFICIENTS - FEMALE MODEL</td>
</tr>
<tr>
<td>10.3.45</td>
<td>-2LL CONTRIBUTION TO FEMALE MODEL</td>
</tr>
<tr>
<td>11.3.46</td>
<td>COMPARISON OF FIT CONTRIBUTIONS - MALE &amp; FEMALE MODELS</td>
</tr>
<tr>
<td>NO.</td>
<td>TABLES</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.2.1</td>
<td>FREQUENCY OF PARTICIPATION IN SPORT DURING THE PREVIOUS 28 DAYS</td>
</tr>
<tr>
<td>4.3.1</td>
<td>ATHLETIC EVENT GROUPS FOR LEVEL 1</td>
</tr>
<tr>
<td>4.3.2</td>
<td>CONDITIONS FOR AWARD - LEVEL 1</td>
</tr>
<tr>
<td>4.3.3</td>
<td>THEORY MANUAL - LEVEL 1</td>
</tr>
<tr>
<td>7.3.1</td>
<td>2 X 2 CLASSIFICATION TABLE (THRESHOLD = DEFAULT = 0.5)</td>
</tr>
<tr>
<td>7.3.2</td>
<td>REVISED CLASSIFICATION TABLE (THRESHOLD 0.7)</td>
</tr>
<tr>
<td>7.4.1</td>
<td>BIRTH AREA &amp; PROBABILITY OF BEING A COACH</td>
</tr>
<tr>
<td>8.2.1</td>
<td>COMPOSITIONS OF THE INITIAL ESTIMATED RESEARCH POPULATION</td>
</tr>
<tr>
<td>8.4.1</td>
<td>QUOTA SAMPLES FOR THE RESEARCH POPULATION</td>
</tr>
<tr>
<td>8.4.2</td>
<td>COMPOSITIONS OF THE ESTIMATED RESEARCH POPULATION</td>
</tr>
<tr>
<td>8.4.3</td>
<td>COMPOSITIONS OF THE VALID INTERVIEWS USED IN THE RESEARCH</td>
</tr>
<tr>
<td>8.5.1</td>
<td>DATA WEIGHTS - ALL 112 CASES USED IN ANALYSIS</td>
</tr>
<tr>
<td>8.5.2</td>
<td>DATA WEIGHTS - REDUCED 110 CASES USED IN ANALYSIS</td>
</tr>
<tr>
<td>8.8.1</td>
<td>REQUESTS FOR INTERVIEW</td>
</tr>
<tr>
<td>8.8.2</td>
<td>INTERVIEW REQUEST CONVERSION RATES</td>
</tr>
<tr>
<td>8.8.3</td>
<td>MEANS OF CONTACT BY GENDER AND BY CLUB</td>
</tr>
<tr>
<td>8.8.4</td>
<td>REASONS FOR INTERVIEW FAILURE BY GENDER AND BY CLUB</td>
</tr>
<tr>
<td>9.2.8.1</td>
<td>LITERACY MEASURES. BY PUBLICATION AND PERSPECTIVE</td>
</tr>
<tr>
<td>9.2.1</td>
<td>CODING FOR AGE, X28</td>
</tr>
<tr>
<td>9.2.2</td>
<td>CODING FOR GENERATION, X29</td>
</tr>
<tr>
<td>10.2.1</td>
<td>SUMMARY OF DESCRIPTIVE &amp; POTENTIAL EXPLANATORY VARIABLES</td>
</tr>
<tr>
<td>10.3.1</td>
<td>LOGIT CALCULATION FOR RESPONSE VARIABLE ‘Car’</td>
</tr>
<tr>
<td>10.5.1</td>
<td>CORRELATION COEFFICIENTS WITH ‘GENDER’</td>
</tr>
<tr>
<td>10.5.2</td>
<td>THESIS MODEL STATISTICS</td>
</tr>
<tr>
<td>10.6.1</td>
<td>dBeta VALUES GREATER THAN 1.00 INDICATING INFLUENTIAL CASES</td>
</tr>
<tr>
<td>10.6.2</td>
<td>VALUES OF dfBeta FOR CASES WITH LARGE VALUES FOR dbeta</td>
</tr>
<tr>
<td>10.7.1</td>
<td>STANDARDISED COEFFICIENTS - THESIS MODEL - BY RANK</td>
</tr>
<tr>
<td>10.7.2</td>
<td>-2LL CONTRIBUTION TO FINAL MODEL - BY STANDARDISED COEFFICIENT RANK</td>
</tr>
<tr>
<td>10.7.3</td>
<td>“Coach” ASSIGNMENT ABILITY BY VARIABLE</td>
</tr>
<tr>
<td>10.8.1</td>
<td>COEFFICIENTS, Exp(B) AND CONFIDENCE LIMITS - THESIS MODEL</td>
</tr>
<tr>
<td>10.8.2</td>
<td>COEFFICIENTS, Exp(B) AND CONFIDENCE LIMITS - REDUCED THESIS MODEL</td>
</tr>
<tr>
<td>10.8.3</td>
<td>COEFFICIENTS, Exp(B) AND CONFIDENCE LIMITS - REDUCED THESIS MODEL (2)</td>
</tr>
<tr>
<td>10.9.1</td>
<td>COMPARISON OF HYPOTHESES WITH THESIS MODEL</td>
</tr>
<tr>
<td>10.9.2</td>
<td>COEFFICIENTS &amp;STANDARD ERRORS IN THESIS MODEL</td>
</tr>
<tr>
<td>10.10.1</td>
<td>COMPARISON OF INCENTIVE/DISINCENTIVE EXPECTATIONS WITH THESIS MODEL</td>
</tr>
<tr>
<td>10.11.1</td>
<td>VARIABLES NOT IN MODEL: COLLINEARITY &amp; SIGNIFICANCE</td>
</tr>
<tr>
<td>10.11.2</td>
<td>AVERAGE LOGITS, BY IN OR OUT OF MODEL &amp; BY GENDER OF RESPONDENT</td>
</tr>
<tr>
<td>11.2.1</td>
<td>DETAILS OF THE MALE MODEL</td>
</tr>
<tr>
<td>11.3.1</td>
<td>DETAILS OF THE FEMALE MODEL</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1. The Research Question

Since at least the 19th century, the voluntary sector of the UK economy has been, and continues to be, a significant provider of social and community services. The voluntary sector provides the majority of the population with much of its sport and leisure activities. This is despite the growth of public and private sector involvement in health and leisure centres during the latter half of the 20th century. (Holt, R, 1989; Cox, RW, 1994; Brailsford, D, 1997; Polley, M, 1998).

These studies show that the voluntary sector caters for sport and leisure activities at local levels through amateur clubs run by and largely financed by the membership income of each club, although the capital facilities are often provided by and hired from either local government or churches or village halls. Up until the 1960’s, the majority of these local clubs largely catered for adults and young people towards the end of their school careers. Sport for younger children was largely supplied by the schools. Changes in academic practices during and since the 1960’s have resulted in sharp declines in the time devoted to sport in schools. From the 1960’s, the network of local sports clubs responded to this vacuum by extending their activities to younger and younger age groups. This increase in membership of young people had to be supervised and managed. This was achieved by recruitment and training of more and more adults, often parents. The majority were and still are unpaid volunteers.

The second half of the 20th century saw rapid advances in scientific knowledge related to physical activity, leading to the establishment of academic disciplines of sports science and sports sociology. Improvements in training methods, sports facilities, equipment, footwear and clothing resulted in ever increasing standards in competition. The general public became more and more aware of these with the growth of mass communication, especially TV.
As a result, the expectations of participants grew, as well as those of the parents. By the 1970’s, it became clear that more formal and structured training of volunteers was required for them to have the extended knowledge required to safely reach the new standards of competition. Both the national governing bodies (NGB’s) of sports in the UK and the government responded. Some NGB’s introduced their own education programmes for both voluntary and professional coaches and officials. The National Coaching Foundation (NCF), now Sports Coach:UK, was set up by the government to provide a modular programme of courses for all sports. Initially, participation in these education programmes was voluntary and many of the qualifications had limited academic or vocational recognition. More recently, the growth of litigation and public concerns about child abuse have led to a growing insistence by NGB’s, local and central government for all practising coaches to be formally qualified. (Houlihan, B, 1991)

At the same time, the UK government has taken an increasing interest in sport and physical recreation. This has arisen from public concern at the lack of success at international level, especially at the 1996 Olympic Games. Also surveys have indicated serious declines in the physical health of the population, especially the young. If allowed to continue, these declines may have future impacts upon health services and upon industrial productivity. Both political parties saw the solution to both these problems, international success and health, as being with provision of better facilities and improvements in coaching. Whilst improved facilities can be provided by public and private sector financing, it is not feasible for these sectors to significantly increase the coaching resource. As a consequence, government is seeking to mobilise the voluntary sector. The strategy not only seeks to increase the number of voluntary coaches and to deploy more of them in the school environment as well as at club level and above (The Sports Council, 1988; Houlihan, B, 1991; Department of National Heritage, 1995).

Volunteering and coaching are more fully discussed in Chapter 2 and 3 and the initial coach education programme for athletics is described in Chapter 4.

INTRODUCTION
To increase recruitment, an understanding must be had of why people volunteer to coach. Whilst some research exists on the motivation of participants in sport, both adult and child, the research into why people volunteer is limited. No research has been found that examines why individuals associated with a specific sport choose to enter the coach education programme and why others choose not to do so. Indeed, very little research exists that discusses both leisure and volunteering (Stebbins, R A, 2000). This mirrors the minimal research that has been undertaken on why adults take part in any form of vocational further education, the basis of Life Long Learning, the subject of current government initiatives.

This research aims to provide an objectively based framework that models the decision process of volunteers making decisions in sport. A generalised model is put forward in Chapter 5, based upon existing research. This model is then developed into a specific model covering the decision to enter a coach education programme. This specific model is based upon the structure, theories and techniques of statistical Logistic Regression. The Logistic regression model is then tested in a specific context of UK track and field athletics (hereafter referred to as 'athletics'), using a sample of adult athletes, parents and existing coaches. Athletics was used as it is a major UK sport involving both sexes and a wide age range and is not geographically dependant. Athletics in the UK also has a relatively long history of offering a formal coach education programme with the vast majority of coaches being unpaid volunteers.

This research is exploratory using a specific sample in order to validate a model that can be adapted to other aspects of athletics or, with further adaptation, to other sports and to other decisions to voluntarily enter adult vocational education. It does not claim to identify a complete set of factors influencing the decision to volunteer as a coach, even within athletics.
This research attempts to show is that there is a substantive and significant set of variables, response variables, that influence the decision to become a coach. The aim is to show that a logistic regression model can be used as a tool for those seeking to design and implement change in the field of coach education courses for adult volunteers in the UK. It is not definitive, in that it does not purport to state a set of causal factors influencing a decision to become or not become a coach. It does not seek to prove statistical significance in terms of prediction.

Logistic regression, in common with any form of modelling, is essentially a descriptive methodology. It does not claim to give a unique or an exact representation of the situation, only one approximation. To quote Collett (1991, p.44):

"Statistical models constructed for response variables are at best an approximation to the manner in which some observable variables depend on the other variables. No statistical model can be claimed to represent the truth and, by the same token, no one model can be termed the correct model. ...... Statistical models are essentially descriptive and, inasmuch as they are based upon experimental or observational data, may be described as empirical models."

The research question posed is:

"DOES THE DECISION TO ENTER THE UK ATHLETICS COACH EDUCATION PROGRAMME CONFORM TO A LOGISTIC REGRESSION MODEL?"

The concept behind such a model is that an individual makes a rational choice between two alternatives when the individual's preference from a weighted set of variables either exceed or fall below the results from the same set using another set of weights. That is:
\[ W_u = \sum a_{u1} X_{u1} + v_u \]  
\[ W_a = \sum a_{a1} X_{a1} + v_a \]

(Equation 1.1.1)

\[ W_u = \sum a_{u2} X_{u2} + v_u \]

(Equation 1.1.2)

where:

- \( W_u \) is the individual's preference for the choice, not coaching
- \( W_a \) is the individual's preference for the alternative choice, coaching
- \( X_{u1} \) and \( X_{a1} \) are the set of variables associated with both choices
- \( a_{u1} \) and \( a_{a1} \) are the coefficients (weights) associated with \( X_{u1} \) for the two choices respectively
- \( v_u \) and \( v_a \) are the approximation errors associated with the estimates \( W_u \) and \( W_a \)

(Collett, D, 1991)

Then the concept is that if \( W_u \) is greater than \( W_a \), then individual i will make the choice not to coach, and that if \( W_u \) is less than \( W_a \), then individual i will make the choice to coach.

How Equations 1.1.1. and 1.1.2. can be turned into a model and how this model can be solved to give estimates of the coefficients are covered in Chapters 6, 7 and 10.
1.2. Background

There are estimated to be 150,000 sports clubs in the UK with 6.5 million members, including 200,000 voluntary coaches, and 500,000 voluntary officials. Together they represent the equivalent of 25,000 full time paid posts. This compares to the estimates of 75,600 people employed in leisure and recreation by local authorities and a further 120,000 employed in the commercial sector. The voluntary sector in sport and recreation in England in 1984/5 put £1,478 million into capital expenditure. This compares to capital expenditure of £226 million by local authorities and of £1,000 million (including revenue expenditure) by the Commercial Sector. In the same period, the voluntary sector spent £14 million in revenue (excluding labour), compared to £752 million (including labour) by local authorities (The Sports Council, 1988, 1996(a)).

UK Sport, formerly The Sports Council, is a government appointed body. In formulating their strategy for the 1990’s, they took account that voluntary sports clubs had a vital role to play. They saw the role of clubs mainly in terms of harnessing enthusiasm and skills, by organising competitions and by helping to improve performance. At the same time, UK Sport thought that these clubs, through their governing bodies, had to improve their marketing. The reason was to compete for membership with the increased provisions provided by local authorities and the commercial sector.

Successive governments have implemented voluntary and sports policies that rely upon forming partnerships between the public and voluntary sectors (Wolfenden Committee, 1978; HMSO, 1995; Department of National Heritage, 1995). These policies imply two assumptions:

- That volunteer numbers and the amounts of time that they give will, at least, continue at existing levels;
- That volunteers will be sufficiently motivated to work in these partnerships.
There are indications that neither of these assumptions may be valid. The 1997 National Survey of Volunteers detected a decline in volunteering, especially in the younger age groups (Smith, J D, 1998). Motivational studies indicate that volunteers, especially those associated with sport, are primarily driven by personal and family desires (Rapoport, R, and Rapoport, R N, 1975; Butt, D S, 1987). Altruism may be only a minor factor (Smith, D H, 1981).

There is growing doubt whether the motivation of volunteers will provide sufficient incentives for them to become increasingly involved in the public sector (Whelan, R, 1996). There is increasing evidence that some volunteers resent working alongside people paid to work in sport. An informal interview in 1997 with Paul Bowen, then Curriculum Manager (Outdoor Education) for Surrey County Youth Services is indicative of the problem. Paul organised outdoor competitive events and had to use volunteers. With a declining budget, he had more need of unpaid assistance. To quote from a discussion, with his permission (Bowen, 1997):

"There is a dichotomy between a declining budget and a decline in the number of volunteers who will work without pay. My feeling is that (a) there is a lot more part time work around, including at weekends, and (b) there is seen to be an inequality between permanent part time staff who we have to pay and volunteers who are not paid and yet do the same job. Envy is beginning to creep in. Volunteers are becoming demotivated. I can not get through to them (the volunteers) the problem (of the budget). In a social context, it seems to me that amateur sport is being eroded away to professional sport."

This resentment may also be fuelled by 1980's culture that the main incentive in life is earning money and that the value of any activity is how much people are willing to pay for it. Sports coaching may be seen in the same light by the generation brought up in this culture. This could also account for the apparent decline in volunteering in the younger age groups found in the 1997 survey (Smith, J D, 1998).
The danger of ill conceived manipulation of the voluntary sector by government has recently been summed up in the Positional Paper of the Volunteerism Commission of the World Leisure and Recreation Association. This statement makes a distinction between “mainstream” and “marginal” volunteering. The latter, “marginal” volunteering, is seen as a voluntary activity where the volunteer feels that their unpaid input is being, in some way, exploited by others, particularly the state, as unpaid labour. This feeling can arise if a degree of coercion or restriction is seen by volunteers to have been introduced to the activity from outside. The statement warns that a change, as perceived by volunteers, from “mainstream” to “marginal” volunteering. The result may be to lessen the impact and enjoyment by the individual of that activity as leisure (Stebbins, R A, 2000).

There are also structural reasons why there may be changes occurring in the participation of people in leisure in the UK. Whilst the average number of hours worked declined during much of the 20th Century, the decline slowed during from 1970 onwards. At the same time, the productivity per hour has increased and the percentage of people in full time occupation declined (unemployment, underemployment, later entry and earlier cessation). During the 1990’s particularly, there has been a marked increase in the number of part time jobs, particularly worked by women. This has led to increases in more than one paid job being undertaken by one or both partners in a household. The effect on sports participation appears to be that social classes D & C2 were participating more in 1993 than in 1983 whilst those in social classes A/B and C1 were participating less (see Table 1.2.1. below).

Table 1.2.1.
FREQUENCY OF PARTICIPATION IN SPORT DURING THE PREVIOUS 28 DAYS
(Central Statistical Office, 1983, and 1993)

<table>
<thead>
<tr>
<th>Social Class:</th>
<th>A/B</th>
<th>C1</th>
<th>C2</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>6.5</td>
<td>5.9</td>
<td>4.2</td>
<td>3.0</td>
</tr>
<tr>
<td>1993</td>
<td>4.9</td>
<td>5.5</td>
<td>5.3</td>
<td>6.0</td>
</tr>
</tbody>
</table>

INTRODUCTION
It has been suggested that two leisure groups are now appearing in the UK:

i. Those that earn a lot, work long hours and under stress who prefer to spend money on buying leisure for comfort rather than for pleasure and who do not wish to become ‘immersed’ in activities other than work. These are mainly the A/B and C1 classes of managerial and professionals, about 35% of the population.

ii. Those who are working shorter hours than ever and who are more likely to be unemployed or underemployed who have more leisure hours but are without the means to pay for leisure, more and more of which has charges set to make a profit. These are mainly the C2 and D classes of office, skilled and unskilled workers.

(Gratton, C, (Ed.), 1996.)

The consequences of these structural changes on voluntary sport may be profound. Traditionally, volunteering has been predominantly a pursuit of higher socio-economic populations. Surveys of sport have indicated that this is true for the running of sports clubs and the provision of coaches (The Sports Council, 1996(a)). The changes affecting these socio-economic groups, together with the increases in both partners being employed and the increases in single parent families, may well reduce the potential number of voluntary sports coaches. Yet the demand for participation may be increasing from socio-economic groups with greater leisure time and from the demand of single parent families for sports clubs to cater for their children.

There are four other factors that could lead to a diminution of the present voluntary coach resource.

i. Pressures from the Sports Council, local authorities and Insurance companies for all coaches to be qualified through formal educational schemes with regular revalidation.

ii. Pressures from the Sports Council and local authorities for all volunteers associated with children, especially coaches, to be vetted in accordance with the Child Protection Act. There are indications that some volunteers perceive this as personal intrusion.

INTRODUCTION
iii. The world wide increases in veteran sport. In athletics, this started with the road running boom of the 1980's. In the 1990's, veterans track & field competitions increased allowing competition to continue beyond the ages of 35 to 40. These trends have since increased with recognition that health is associated with exercise. Such participation may be more attractive to the individual than becoming a coach. Veteran sport adds to the demands for coaching and may also reduce the number of people becoming coaches.

iv. Increases in the number of paid jobs specifying a coaching qualification. Younger people, in particular, may choose to gain a coaching qualification to enhance their career rather than to become a volunteer.

1.3. Research Basis

Chapter 2 reviews the research on volunteering and the theories developed, especially the motivational aspects and the variety found within the voluntary sector. Chapter 3 reviews the research on sport as a social activity and it’s relations with government. From this, the research develops a generalised model of decision making in sport based upon Social Exchange Theory. This Theory is discussed more fully in Chapter 5. Exchange Theory can be traced back to the Adam Smith’s work on the division of labour of the 1770’s (Smith, A, 1991; Marshall, G, 1998). An adaptation by sociologists became known as Social Exchange Theory (Homans, C, 1961; Bottomore, T, & Nisbet, R, 1979; Coleman, JS, 1994). A concept of decision thresholds is then introduced, based upon a Recreational Adoption model (Brandenburg, J, and others, 1982). This generalised model is then further adapted to represent the specific decision to enter a coach education programme. Chapter 4 describes the coach education programme used in athletics in the UK until 2000, when it was substantially revised. The model is then modified for athletics and tested against a sample of adults associated with athletics either as adult competitors or as parents of child participants or as coaches who have entered the coach education programme. The model is formulated to predict whether or not an individual from the sample population is likely to decide to enter the coach education programme.
"Decision" considered this way is dichotomous in that it can only take one of two states: "Enter" or "Not Enter". For definition and to adhere to independence of the dichotomy, individuals who have not considered entry are classified as "Not Enter". Statistically, the variable "Decision" is a Binary variable. The two states, "Enter" and "Not Enter", are arbitrarily assigned the values of "1" and "0" respectively.

Use of a binary target variable immediately restricts the types of statistical models that can be used. Traditional statistical models require the target variable to be able to take any value along a continuous scale. Furthermore, some of the influencing factors that must be considered are also binary, notably gender. This fact further restricts the model type, logistic regression being the main one. Logistic regression is a specific technique within the class of generalised linear models (GLM) (Hutcheson, G, and Sofroniou, N, 1999). GLM and logistic regression are discussed in Chapters 6 and 7.

A basic assumption of any form of statistical regression modelling is that all important variables are included in the model. Existing research has been examined in order to draw up a list of variables that have either found to be important in previous statistical research on the subject or have been hypothesised in research theories. As very little research has been published on statistical modelling of the sports coach decision process, it was necessary to examine the published research on volunteering and particularly on sports coaching. The literature surveys undertaken are reported in Chapters 2 and 3.

Ideally, modelling of the decision to become a sports coach should be tested to cover all adults associated with sport in the UK. This would mean sampling a multimillion population of all adult members of all sports clubs in the UK, together with the parents of all juvenile members, a total population probably in excess of 5 million individuals associated with 150,000 clubs in over 100 different sports. The problems, costs and time required to negotiate access to a sufficiently large random sample of membership lists needed to contact the population would be beyond the resources of this research. Furthermore, decisions on coaching in one sport may be influenced by
different factors to another sport as a result of differences in the roles of coaches in each sport and the characteristics of the different coach education programmes. The research recognises that any framework model will require tailoring with the characteristics of an individual sport for it to be a working tool for that sport.

A single sport, athletics, was chosen as the NGB have been running a national coach education programme for over 25 years without major changes and were willing to participate in the research. The research was further restricted to two similar clubs in Surrey. This restriction was determined partially by available time and funding resources and partially by the need to minimise the number of variables that were considered. The number of variables influences the minimum sample size required to test for significance. Sampling and sample size are considered in Chapter 8.

This restricted sample is justified for the research as the objective is to show that a substantively significantly model of the decision making can be achieved using the framework of logistic regression. The research does not purport to identify every significant variable, even for the restricted sample, nor does it aim to accurately rank the effectiveness of the variables included in the model. The research is not designed to prove causality of the factors measured by the variables to the decision, only that there is some form of association between a variable and the observed outcome of the decision for the population used in the research.

The key stages which this research went through are shown in Diagram 1.3.1. The researcher is interested in the broad question of ‘why do people volunteer?’ For the practical purpose of a specific research programme, this question was narrowed to coaching, to athletics and to the UK. The first steps were to carry out literature searches into volunteering and sports coaching and to decide on suitable definitions of the key words of ‘volunteer’, ‘sport’ and ‘coach’ and, as the question implied it, ‘decision’.

INTRODUCTION
KEY STAGES & PROCESS OF RESEARCH

BROAD QUESTION: “Why do individuals volunteer?” Or “Why do individuals not volunteer?”

NARROWER QUESTION:

“Why do individuals volunteer to coach athletics in the UK?”

DEFINITIONS

LITERATURE SEARCH

DECISION THEORY

EXPECTANCY THEORY

SOCIAL EXCHANGE THEORY

BRANDENBURGH THEORY

RAPOPORT Key Events

MINTERN Thresholds

SPORT MODEL

SAMPLING VARIABLES

“Variables”

“Concept”

“Methodology”

GEN. LINEAR MODELS

LOGISTIC REGRESSION

SAMPLING POPULATION

BINARY MODELS

MODEL TRIANGULATION PROFILES

DATA SAMPLE

RESTRICTIONS, LIMITATIONS ANSWERS FURTHER QUESTIONS

Comparisons Sponsor Rewards

QUALITATIVE INFORMATION

QUANTITATIVE INFORMATION

RESEARCH QUESTION:

“Does the decision to enter the UK Athletics Coach Education Programme conform to a Logistic Regression Model?”

INTRODUCTION
Whilst Diagram 1.3.1. indicates a set of sequential stages, in practice there was a great deal of iteration both between stages and backwards and forwards.

The left hand side of Diagram 1.3.1. shows how the literature search led to consideration of the broad area of Decision Theory and, eventually, allowed a conceptual sports model to be developed. In parallel with this, a set of potential variables was created from consideration of the findings of previous researchers. The results of the literature search are reported in Chapters 2, 3 and 4. The development of the concepts is described in Chapter 5.

The right hand side of Diagram 1.3.1. shows how consideration of definitions led to the examination of binary statistical models and the selection of logistic regression. In a parallel process, the sampling population was defined. This included consideration of the research resources and time available, together with the access to sources of information, in particular name and address lists. These matters are dealt with in Chapters 6, 7 and 8.

The bottom section of Diagram 1.3.1. shows the relationships of the processes involved with answering the research question. Again, iteration took place. Data was collected for the sampling variables. This process required the development and testing of an interviewing process. The collection of the data required the development of a sampling process. Both these processes had to respect the requirements of the chosen methodology. Whilst the use of statistical methods required quantitative data, a decision was taken to also collect some qualitative data. This would allow some triangulation of the statistical modelling. At the same time, it would produce some valuable information for the governing body of athletics, the participating clubs and other researchers. It also made some of the respondents feel much more involved in the research. These are reported in Chapters 8, 9 and 10.
The last line of Diagram 1.3.1. show the results of the research. The research question was answered, but some restrictions and limitations of the methodology and underlying concepts were revealed. Further questions for future research were identified. The qualitative information did allow some triangulation. It also allowed comparisons with previous research findings. Hopefully, it produced information of value to all those bodies and individuals who made the research possible. These aspects are reported in Chapters 11, 12 and 13.
2. VOLUNTEERING

2.1. Introduction

There are three main reasons why an individual may become associated with sport: recreational, social and employment. This research primarily concerns individuals associated with sport for recreational or social reasons. Such people will be termed 'volunteers'. Not only are volunteers and the voluntary sector at the heart of sport in the UK, they are an essential part of a healthy economy and a healthy society. Very little research has been conducted on the organisation of voluntary bodies (Harris, M, 2000) or on why individuals volunteer in sport.

2.2. Definition

There have been many definitions made of 'volunteer'. Most of them centre on whether or not individuals receive any payment. This is sometimes relaxed to “not receiving payment other than the reimbursement of expenses incurred on behalf of the organisation”. This can include paying for a ‘volunteer’ to go on a training or educational course, including coach education. Over the past twenty years, a growing number of organisations now pay their ‘volunteers’ some form of honoraria or pay them for any professional services rendered to the organisation. In the case of voluntary coaches, payments vary from exemption from membership fees to nominal payments for coaching services. Under an amendment to the Charities Act in 1999, sports clubs can now be registered as a charity and are legally entitled to pay volunteers in the ways outlined above. The only criterion applied is that members (volunteers) must not directly benefit from membership of the charity. That does not preclude a charity employing people, although it does preclude those people being members of the charity.

Definition is particularly complex in sport with the growing number of 'professionals'. The sports 'profession' followed by a member of a sports club can be directly
associated with the sport of the club, either as a player or an administrator or a coach or manager. This has become increasingly common as sport has moved from a purely ‘amateur’ basis to a ‘professional’ basis, especially at international levels. Even allowing for this, members of a sports club may be employed professionally outside the club as teachers, especially as physical education professionals; or as local government employees, especially in leisure centres and as sports development officers; or in private sector health and fitness clubs. In these cases, it is often difficult to determine whether their association with the sports club is purely ‘voluntarily’ or is motivated, wholly or partially, by their profession. This is particularly highlighted by the possibility of younger people taking a coaching qualification in order to enhance their CV for entry to Higher Education or for a job application.

Alternatively, the characteristic of the organisation can be used as the determinant of ‘voluntary’. This has led to the concept of the non-profit sector, rather than purely a voluntary sector. The non-profit sector lies between the state sector and the for-profit sector. The non-profit sector has basic characteristics:

- provide useful public or semi-public goods or service and serve a specific public purpose;
- do not distribute residual income as profit;
- voluntary, in the sense of being created, maintained and terminated by voluntary decision;
- based upon a value rationality, a deep rooted set of values that motivate and guide the organisation.

(Anheier, HK, 1990).

Throughout this research, the terms ‘voluntary sector’ and ‘non-profit sector’ will be used interchangeably, using the term most suitable for the argument being developed.

David Smith (1981, p.22) defined ‘volunteer’ on the motivations:
"an individual engaging in behaviour that is not bio-socially determined (e.g. eating, sleeping), nor economically necessitated (e.g. paid work, housework, repair work), not socio-politically compelled (e.g. paying taxes, clothing oneself), but rather that it is essentially (primarily) motivated by the expectation of psychic benefits of some kind as a result of activities that have a market value greater than any remuneration received for such actives."

This definition has been adopted for this research. It avoids many of the dilemmas discussed above. It is applicable both to individuals and to organisations. It is compatible with the various definitions used by UK government departments and the UK Volunteer Centre. The key section of the definition for a volunteer in sport is ‘.....essentially (primarily) motivated by the expectation of psychic benefits of some kind as a result of activities ....’ In other words, if an individual is a member of a sports club primarily to personally benefit from participation in the sport, then that individual can be seen as a ‘volunteer’ when assisting the club.

Before examining motivations of volunteers related to sport, it is useful to consider how these motivations have arisen through the history of society. First, volunteering as an activity within society is considered.

2.3. Volunteering and Society

Over the past 200 years, social provision has gradually moved from the voluntary sector (The church, guilds, charities) to the state. In England this process reached its climax after world war II with the implementation of the welfare state. In the 1950’s both local and central governments began to encourage voluntary organisations and, at the same time, to try and co-ordinate their activities. The concept of a pluralistic society was borne (Wolfenden Committee, 1978). Wolfenden pointed out that freedom for individuals to intervene, or to abstain, was a crucial characteristic of the voluntary sector. He drew attention to the peculiar British characteristic of preferring to support specific activities. His Report recommended partnerships between the
voluntary, public and private sectors. This followed the ideas of Richard Crossman. Crossman was convinced that volunteers were better able to perform certain tasks better than professionals. He advocated co-operation between professionals and volunteers and government to improve performance (Crossman, R., 1973). A view developed that volunteers filled the gaps left by professionals and that the voluntary sector maximised use of the talents of the population. It was argued that this was a right of the individual and demonstrated democracy (Lyon, A., 1975).

But Wolfenden warned that a uniform overall pattern for all to conform to was not suitable due to the wide variations of the need and of the desire to volunteer. Every pattern of working is designed to maximise the chances of achieving the organisational aims, as well as obtaining a mutually agreeable balance between the aims of the workers and the aims of the suppliers and customers. Whilst the aims of a voluntary organisation may be a sub-set of the aims of a public sector organisation, it is manifestly obvious that the aims of their workers have significant differences in that volunteers do not seek to be paid a living wage whilst workers seek jobs to gain a living wage.

Furthermore, the bonds between a volunteer and a voluntary organisation are, in some respects, weaker than those between a worker and an employer. Both the volunteer and the voluntary organisation can sever their bonds at any time without any substantial damage to the individual. The volunteer can devote the time released to other ends and will not suffer financially. The worker is contractually and legally bound. The worker may find difficulty in getting employed again and may suffer financially, along with any dependants. Especially in times of high unemployment or rising inflation, the aims of workers will increasingly include safeguarding their employment.

A volunteer seeks voluntary organisations to gain perceived benefits to themselves or their dependants, some of which benefit may be personal satisfaction in giving service to others. On the whole, these benefits and satisfactions are intangible and

VOLUNTEERING
psychological or sociological. Furthermore, required benefits may change with time, For example, the motivation may arise from the desire to take part in, or for a dependant child to take part in, a physical activity. As the person grows older or the child leaves home, the benefit and satisfaction from a particular physical activity may decline. In contrast, every worker seeks a job to gain tangible benefits, in particular a regular income. This does not change with time, even in retirement.

Consequently, any successful form of co-operation between public and voluntary organisations would have to reconcile these differences and any pattern would have to be sufficiently flexible, on both sides, to accommodate changes. Such patterns have become known as partnerships.

During the 1970’s the concept of partnership began to be implemented and reflected in government policies. The balance of delivery of welfare services by the public and the private sectors was reviewed as the economics of the welfare state became a concern to government.

The majority of the research on volunteering had been in the United States and, to a lesser extent, Germany, the United Kingdom and France. In 1986, only about 30% of research studies concerned specific areas of activity. The majority were general or covering a group of related activities. This was probably due to the dominance of government or quasi-government funding. In the United Kingdom, management issues were a dominant subject for research. Historical, legal, fiscal and political factors have influenced how Non-profit (voluntary) organisations have developed. Consequently, studies done in one country have limited applicability to other countries, particularly as no general theories have been developed (Anheir, H K, 1990).

In the UK during the 1980’s and 1990’s, government policies increasingly placed greater responsibilities on voluntary organisations for health and social service delivery. The work of local voluntary organisations changed as a result. A review of
research on the subject found great variety in motivation of volunteers and of reaction to change (Billis, D, and Harris, M, 1992). In discussing the volunteers, this review states (p.216):

‘But they share a common belief in the worth of the work that their group does for others, and they are “not as strongly guided by an abstract sense of civic duty as by a desire to help others”’ (Chambre, S, 1989).

In other words, volunteers are motivated more by an “associational” than a “state” welfare ideology.’ This is particularly applicable to voluntary membership of sports clubs.

The Billis review also found that as agencies grew and became more influenced by government and central bodies, volunteers became over stretched. The overstretching may have been as a direct result of additional work or it may have been due to conflicts of interest between the bodies and the volunteers. Whatever the cause, increasing difficulties in attracting and retaining new members followed, possibly because potential recruits were deterred, although it could have coincided with changes in social views on volunteering. The main deterrents to volunteering appeared to be the amount of work, responsibilities, training and commitments entailed. Studies are cited showing how existing volunteers began to feel “devalued” with the introduction of financial inducements and of paid staff into their area of activity. Whilst these studies concerned health related voluntary organisations, the observations may have much more widespread implications. One example of the problems that can arise was given in Section 1.2 relating to the Surrey Youth Service.

The importance of the voluntary sector to government is reflected by the inclusion of questions about volunteering in the General Household Survey from 1981 and by the commissioning of the National Survey of Volunteering in the same year and subsequent updates. Until 1997, successive surveys indicated that volunteering was on the increase, both in terms of the number of volunteers and of the amount of time
each volunteer gave. But the Report of the 1997 National Survey detected signs of decline, especially amongst the younger age groups:

“A worrying picture emerges in relation to young people. In the case of the 18-24 and 25-34 year olds levels of formal participation are back to their 1981 levels; the increase of 1981 to 1991 appears to have been wiped out. It is not only overall levels that are down. The time given by existing young formal volunteers has plummeted from an average per week of 2.7 hours in 1991 to an average of 0.7 in 1997”. (Smith, J D, 1998).

In 2000, it has been estimated that the financial contribution to the UK economy of the voluntary sector was £12.8 billion a year, or nearly 2% of the Gross National Product (Longley, C, 2000; citing Harris, M, 2000)

One theme that seems to occur very frequently in this brief historical survey is the concept of ‘altruism’. To many, it is the concept of ‘altruism’ that seem to characterise activities that are considered to be ‘voluntary’. This concept and it’s relations with ‘volunteering’ is next discussed.

2.4. Altruism

David Smith (1981, p.23) defines “altruism” as:

“An aspect of human motivation that is present to the degree that the individual derives intrinsic satisfaction or psychic rewards from attempting to optimise the intrinsic satisfaction of one or more other persons without the conscious expectation of participating in an exchange relationship whereby these ‘others’ would be obligated to make similar/related satisfaction optimisation efforts in return.”

The Wolfenden Report held the view that the motives for volunteering were partially altruistic and partially for self advancement. The inclusion of altruism as a motivation
to volunteer was reinforced at that time by contemporary research (Titmuss, R M, 1975; Anderson, J C, & Moore, L F, 1978; Leaf, D, 1978). It is still the assumption of government that altruism is a significant motivator (HMS O, 1995).

But David Smith (1981) questions the value of some of the earlier research on the basis that it was based upon a question such as "Why do you volunteer?" which may only elicit socio-culturally acceptable answers. Other research questions the importance of altruism.

A study of leadership in neighbourhood organisations concluded that altruism was not the essence of volunteering, rather contributions in the form of time or money or materials. These allowed the achievement of some perceived aim shared by the individual and the organisation. (Rich, R C, 1980).

A more realistic view is that an organisation may or may not be altruistic and its members will display altruism to a greater or lesser extent. An organisation is only altruistic if its one, and only, purpose and goal are, in reality and not just for publication, directed to the allocation of its resources toward the optimisation of the satisfaction of non-members without expectation of a quid pro-quo exchange of any sort. On this argument, most organisations are probably not altruistic and nor are many members. Only some members may be primarily altruistic, e.g. fund raisers. Certainly, volunteers in sports clubs can not be seen as ‘altruistic’ under this definition as are primarily aimed at the satisfaction of themselves and other members. Any activities aimed at satisfying non-members are usually to recruit new members. This includes assisting local schools with sport.

A study of the motivations and expectations of volunteers in a town in the United States suggested that altruism may be an initial motivation, but the experience of volunteering may cause other motives to appear and become predominant. Throughout the study enlightened self interest was a continual theme (Phillips, M, 1982).
A Theory of the Voluntary Sector was offered in 1988 by Weisbrod. This was used to investigate the role of sports clubs in the changing market for sport (Deckers, P., and Gratton, C, 1995). They concluded that motivations to participate were predominantly those of self interest and that clubs existed for the benefits of the members, not the collective benefit of the sport. Altruism appears to be a minor factor. They also found that each club was closely identified with an individual community and that parents were the main source of voluntary help for many activities in each club. The significance of their findings is that any meaningful attempt to identify motivations must take into account the environmental and membership differences between clubs, even in the same sport.

Robert Stebbins (1997, p.2) draws attention to the importance of altruism and self interestedness in volunteering in leisure and the intricate interplay between them:

"An important assumption ----is that volunteering, among other things, is primarily creative, society-building activity, which nevertheless loses this quality when it is foisted on altruistic citizens by the public or private sector as a money-saving strategy."

These indicate that altruism may not necessarily be a dominant aspect of voluntary work. Furthermore, its significance may change over time and become less dominant. The alternatives are that volunteers are motivated by concerns about self, work, family and their personal environment and interests. If such non-altruistic motivations are predominant, then there must be some doubt as to whether volunteers will be willing to extend their activities into areas that they perceive to be outside their personal areas of concern. The exception may be if there are sufficiently attractive inducements offered, including money and symbols of public recognition.

Even if a sense of 'altruism' is involved in volunteering, it does not explain why an individual affiliate to one voluntary organisation rather than to another. Consideration of the reasons for affiliation to a specific club may reveal more about motivations.
Recruitment by voluntary bodies has been well researched. It is generally accepted that volunteers affiliate for the following reasons:

- most volunteers became involved through friends and acquaintances who were already volunteers;
- different people were involved in different activities;
- women volunteered more than men, especially in statutory organisations;
- manual workers were more likely to help neighbours than to volunteer;
- unemployed and recently retired were less likely to volunteer than those employed;
- those without children were less likely to volunteer than those with children.


Affiliation to a voluntary organisation has been shown to be a multidimensional process. A study of volunteering amongst US college students using logistic regression (Schanning, K F, 1999) found that three sets of factors were significant: External contextual, relating to parental attitudes, community activity and education. Internal contextual, relating to attitudes towards volunteering and the community. Situational, relating to car ownership, number of friends and organisational membership.

The research concluded that multi-variate model was needed to explore this form of social exchange.

The timing of the decision to become a volunteer may be influenced by external factors. Using logistic regression and other statistical techniques, a study of 180 adults aged 25 or over who entered a graduate education program found direct effects of age and gender, but not race, through career and family transitions. The study also
found that the individual transitions in themselves were not significant, only when they were considered as groups. The two transitional groups comprised:

i. Career: retirement, career change, promotion, demotion, job termination

ii. Family: death, marriage, separation, divorce, childbirth, empty nest

(Willis, P L, 1998).

A study of over 3,000 adults over 18 in the US showed that not only was the decision to volunteer multi-dimensional, the factors associated with the decision varied with the type of voluntary association involved. The study looked at 5 forms of voluntary association and selected 9 social and demographic variables from literature relevant to participation. The 9 independent variables were:

i. Education: 0-8, 9-11, 12, 13+ years of full time education.

ii. Race: white or black

iii. Gender: Binary

iv. Age: 18-29, 30-54, 55-64, 65-74, 75+ years old

v. Marital status: single, married, separated/divorced, widowed

vi. Head of household: head, not head

vii. Number of children under 18 years of age residing at home: 0, 1, 2+

viii. Size of community: 100,000+, 2,500-99,999, under 2,500.

ix. Region of residence: West, East, Midwest, South (USA)

The conclusion was that affiliation was not a unidimensional process and that memberships in different types of associations were not interchangeable. For
recreational activities, age race and region of residence were the significant variables (Williams, J A, and Ortega, S T, 1986).

Incentives can be an important factor in both recruiting and retaining volunteers. The amount and nature of the necessary incentive have to be related to the demands placed upon the volunteers by the organisation. The corollaries are that volunteers must perceive that the incentives are suitable before they will join and that organisations must review incentives as they or their tasks change and grow. Failure on either count is likely to lead to reductions in both the quantity and quality of voluntary effort (Clark, R B, and Wilson, J Q, 1961). Volunteers may vote with their feet.

In the 1990’s research began to show declines in membership of sports clubs in the Netherlands, Germany & Denmark (Heinemann, K, 1993; Deckers, P, and Gratton, C, 1995). As a result, they investigated the reasons why members left sports clubs and why people were not joining sports clubs. They concluded that membership decline was through decreasing recruitment and through increasing resignation. Fewer people, especially the young, were joining because they could access better training and coaching resources at public and private sector sports and leisure centres than at voluntary sector clubs. People were resigning because there were more opportunities in the sports and leisure centres for recreation and leisure without the duties of volunteering. The warning was that increased government involvement with sport could lead to declines in voluntary provision. These findings illustrate the necessity for taking into account both the structures of individual clubs and the environment in which they are situated.

Whilst these studies indicate possible reasons for choice, they do not constitute a theory. One dominant theory, Expectancy Theory, centres on the expectations that an individual considers in making a decision to join a voluntary organisation and what that individual will expect out of affiliation to be satisfied with the choice. This is now reviewed in Section 2.6.
2.6. Expectancy Theory

Much of the research on the reasons why people became involved in an organisation has been based upon the concept of Expectancy Theory. The concept is that an individual engages in an activity in order to obtain a reward: they expect something for their contributions. Research specifically on volunteer involvement indicates different patterns of motivating factors, the patterns being associated with the variables of age, gender, employment, education and social class (Smith, J, 1972, cited in Dowling, D, Potrac, R P, and Jones, R, 1997.). Other studies show a variety of motives and that a decision to volunteer is not a result of a conscious analysis of choices or opportunities.

Motives included: service to society; self actualisation; need for power; emotional associations; autonomy; need for mutual support (Schindler-Rainmann, and Lippett, 1971); self development; social interaction (both to confirm self image) (Johnson, R H, 1973); self esteem; altruism; gaining understanding of an occupation and obtaining 'real world' experience (for a student) (Anderson, J C, and Moore, L F, 1978). The latter not only found the variety of motives, but demonstrated that patterns of motivation varied between populations when they were distinguished by gender, age, employment and social class. They also concluded that age, sex, employment, education and social class were important variables associated with patterns of motivation to volunteer.

The motivations of an individual are dynamic and can change from entry into and during their career in an organisation. For example, voluntary blood donors became committed over time and their motives then included and became more centred on personal goals, such as how often and how much blood they gave (Kessler, R C, 1975). This would call into question policies based upon group goals. Kessler also observed a cycle, about 10 years long, through which a group appeared to flourish.
and then decline. The cycle was related to members retiring and, eventually, having to be replaced by new members who bring in a different set of motivations.

From Expectancy Theory it has been hypothesised that people volunteer to obtain rewards that they are unable to attain through their work. However, other research suggests that people who were socially involved in their work tended to choose leisure activities with minimal social content, i.e. a contrast (Meissner, M, 1971). From this finding, they suggested that these could be explained by hypothesising that people either perceived that they had control over their environment or fate (an internal locus of control) or did not. They predicted that those people with an internal locus of control would become more involved with volunteering than those without this control where employment was not rewarding.

It was then hypothesised that the act of free will involved in working voluntary created a different form of relationship to that at work, which is based upon compliance (Gidron, B, and Baerward, P, 1983). They suggested that volunteers found job satisfaction if the voluntary work gave them some of the following:

- challenge, interest;
- made use of skills or knowledge;
- allowed independence;
- required responsibility;
- hours & location convenient;
- work not hampered by organisation or other factors;

Other studies have shown other factors:

- supervisory support;
- peer contact;
- recognition by others.

Gidron & Baerward further suggest that the relationship of the volunteer, the individual, to the activity is more important than the interpersonal relations.
Furthermore they suggest that, because of the limited time available, volunteers only want to spend time on the central task, not on tasks perceived as being irrelevant.

Much of the research has looked at volunteering as a whole. More recently, the sociology of leisure has received an increasing amount of attention in it’s own right. Some of this has considered the role of the volunteer. In the Western world, most people are working fewer hours now than in the first half of the 20th Century and earning more. But this has caused the emergence and rapid rise of consumerism. People now spend more and more time seeking, buying, using and maintaining material goods and services. They seek to save time by engaging in activities that give rapid returns and are increasingly willing to ‘pay’ others for leisure time. This has been termed ‘time famine’ (Linder, S B, 1970). He suggested that one result was that fewer people volunteered.

Expectancy Theory would indicate that voluntary coaches have a variety of motivations and that these may change as they progress as coaches. From the work of Gidron & Baerward, it can be expected that coaches may, at some time in their progress, be more interested in the act of coaching for their club and might not be interested in coaching outside (e.g. at schools). Perhaps with more serious implications due to the nature of coaching, some voluntary coaches may not see interpersonal relations as being important. Furthermore, the work of Linder would suggest that growing consumerism may be expected to bring about a decline in voluntary resources.

What this review highlights is that the act of volunteering is associated with a large set of variables and that there is not a small subset of these variables that predominate. Furthermore, some of the variables are categorical and others, particularly attitudes, can only be measured on artificial scales. It would also seem that the predominant set of variables may differ from club to club according to the nature and structure of the club. It is not surprising that some research is inconclusive, especially if relies on simple parametric analysis involving a limited set of variables and sampling over a
variety of organisations and sports. It reinforces the basis of this research in limiting the population and using Logistic Regression that allows the use of categorical and continuous variables.

Having considered volunteering in general terms, the peculiarity of sport within the voluntary sector is examined in the next Chapter. These will more sharply focus on the factors that may influence individual volunteers in becoming coaches in sport.
3. SPORTS COACHING

3.1. Introduction

Whilst many sports clubs in the UK are voluntary organisations, they have certain characteristics that distinguish them from other organisations in the voluntary sector (Whelan, R, 1996; Harris, M, 2000). Until changes to the Charities Act in 2000, one of the most important factors has been that a sports club could not be registered as a charity. This meant that, until 2000, every sports club had to rely on raising its revenue from amongst its supporters and from grant aid from Local Authorities and the State. As fund raising is not one of the main motivators for people joining a sports club, such clubs try to be self sufficient in order to minimise costs and the need to seek outside funds. It is probably this reason that has led many sports to use voluntary coaches recruited from their own members, rather than employing professional coaches.

Another feature of sport as a voluntary activity is that a sport requires Rules of Competition at least agreed by a large enough number of clubs to allow competition between them. For a sport to be significant, the Rules have to be at least national, preferably international. This can only be achieved by the existence of a National Governing Body. Whilst there are numerous non sporting voluntary organisations, the vast majority of non sporting organisations are small and not organised nationally. The existence of a NGB tends to lead to two profound effects:

- Governments prefer to talk to one body rather than a myriad of small clubs and to shape policy as a result. The motivations of volunteers serving on NGB's are likely to differ from those of the volunteers in the clubs they represent. This will effect the information on which government acts. In the case of Sport, government has created the Sports Council, now Sport England, to further ease consultation.
• A NGB will tend to be used to, at first, suggest and, later, impose standardised systems, including coaching. This can be seen as reducing the freedom of the local volunteers, this freedom being part of the motivation to volunteer. The consequence is that tensions are produced between local clubs and the NGB. This tendency to impose can now be seen with UK Sport (previously known as the Sports Council), particularly since 1997 when it was appointed as a National Lottery Distributor.

'Sport' has to be defined from the more vague terms of 'recreation', 'exercise' and 'leisure'. 'Coach' has to be defined from related terms of 'instructor', 'manager' and 'teacher'. Both have to be appreciated in the context of their history in the UK in order to appreciate their current structures and their appeal to the individual. Decisions made about coaching will be conditioned both by facts and by perceptions. Perceptions are conditioned by personal experience, including history.

The historic evolution of each sport and the costs required to play them have led to differing socio-economic patterns of people associated with each sport. The characteristics of each sport (team, individual, contact, strength, endurance, etc.) also led to differing motivational, gender, age and physical characteristics of those associated with each sport.

To control these variables, one sport, athletics, has been chosen for this research. However, athletics in the UK:

• Is one of the largest sports, in terms of participation. This allows a large population from which to sample.
• Is one of the most successful, in terms of international medals. This has caused governments to give more notice to it.
• Has a well established coach education system, commencing in the 1970's. This results in a relatively stable coach population on which to research.
• Is still, predominantly, voluntary and amateur. The population available to be sampled, therefore, contains few non-volunteers as defined for this research.
3.2. **Definitions**

3.2.1. **SPORT.**

For the purposes of this research ‘sport’ is defined as:

> "any one of the activities recognised by the Sports Council".

(The Sports Council, 1996(b)).

The Sports Council Guidelines for recognition of a sport require a written constitution with a non-profit distributing governing body that has operated for at least three years with regular participation of at least 5,000 individuals. It expects each activity to have an organisation for introducing skills; to have leadership or coaching awards and to have both a domestic and an international competitive structure.

These requirements place recognised sports in the non-profit or voluntary sector.

3.2.2. **COACH**

Little and others (1983) traced the word ‘coach’ to being used in universities for a private tutor in 1848 and then being transferred in 1885 to training athletes. ‘The National Occupational Classification of Canada Category 5232 - Coaches’ (Human Resource Development, 1997) uses a definition close to the 1885 meaning, extending it to preparation and training.

The Classification expands this definition by listing the main duties of a coach as some or all of:

- “Identifying strengths and weaknesses of athletes or teams;”
- “Plan, develop and implement training and practice sessions;”
- “Nature and develop athletes’ potential skills and abilities;”
• Develop, plan and co-ordinate competitive schedules and programs;
• Motivate and prepare athletes or teams for competitive events or games;
• Formulate competitive strategy, develop game plans and direct athletes and players during games or athletic events;
• Analyse and evaluate athletes' or team's performance”.

This classification is meant to apply to those in employment or self employment, i.e. professional coaches. It includes ‘sports scouts’ who are responsible for identifying and recruiting athletes for professional teams.

No national or international body of sport, including the UK Sports Council and the UK National Coaching Foundation, were able to supply an official definition of ‘coach’, other than the circular arguments such as that of the National Coaching Foundation:

“Any person who holds a current certificate of coaching issued by a governing body of their sport and who operates within the guidelines of the appropriate level of their coaching award”.

However, informal discussions with officials of a number of athletic clubs indicate that a number of people coach regularly who have never held a coaching qualification. Some of them are teachers who have physical education responsibilities in their work. Others are deemed by their clubs to have sufficient knowledge. These discussions are confirmed by comparing the Surrey County Athletic Association Handbook with the Coach Register of the NGB. The entries in the Handbook are from officially submitted lists from each club. 25.7% of coaches listed in the Handbook are not on the Register, although some of this discrepancy may be due to deficiencies in the compilation of the Register.

For the purposes of this research, the definition of the Canadian Classification, together with its list of duties, has been used. Distinction will be made between ‘Registered’ and ‘Unregistered’ Coaches, the former being those on the NGB register.
Unless otherwise stated, the undifferentiated term ‘Coach’ will imply ‘Registered Coach’.

3.3. Sport and Volunteering

In the United Kingdom, the concept of the amateur and volunteer has deep roots with the many of the governing bodies of sport that were formed in the 19th century and early 20th century. Frequently their rules specifically excluded from membership and from competition anybody who received any form of payment. Indeed some amateur sports associations (e.g. the Amateur Rowing Association) also excluded individuals who were employed in any employment deemed by that association to be manual labour. Professionalism was catered for, usually by having separate governing bodies (e.g. Boxing) and forms of competition (e.g. Rugby Football), the amateur and professional associations of any one sport being mutually exclusive memberships. Prize money in competitions was either forbidden or restricted by the rules. Rules might also exclude the exchange of money by betting (e.g. the Amateur Athletics Association). Exceptions to the mixing of amateurs and professionals were the established sports of Association Football (‘Soccer’), Cricket and Golf. The governing bodies involved allowed membership, clubs and teams with both amateurs and professionals. Even here, the amateurs were very much in control of the sport and were referred to as ‘Gentlemen’. However, where professional participants were involved, early recognition of the economic value of these professionals was made by allowing clubs to charge transfer fees.

Amateur competitors employed professional coaches and advisors. This practice was deprecated by many of the amateur officials. For example, the Rules of London Rowing Club, established in 1847, allowed payment of a professional Waterman in a crew for training. The payment was a fine levied on a crew member who was late for an outing. This Rule existed until the 1960's. At the beginning of the 21st Century, athletics in the UK, very few coaches are paid salaries, fees or expenses. Perversely,
amateur officials (e.g. time keepers, judges and starters) have, by tradition, been paid their expenses and some administrators have been paid fees.

Further moves away from the original amateur, voluntary, unpaid base of sport has occurred associated with:

- increases in the provision of public sector sports facilities, especially leisure centres;
- reductions of subsidies and ‘privatisation’ of local authority sports provisions;
- increased public awareness of health and body shape;
- increases in the disposable incomes available for leisure, allied with increased availability of cars for private use;
- changes in the employment, work and status of teachers, especially those involved with games and physical education;
- strategies of the government appointed Sports Council.

Despite all these changes, sport in the UK at the end of the 20\textsuperscript{th} century was still highly dependent upon volunteers, however these are defined (The Sports Council, 1996(a)). This was confirmed and quantified by a major national survey conducted for the Sports Council by the Leisure Industries Research Centre (The Sports Council, 1996(a)). This has led to a series of government initiatives aimed to increase the number of volunteers in sport and to increase their effectiveness. The recruitment initiatives involve financial support from the Exchequer and the National Lottery, largely distributed by local authorities. The effectiveness initiatives are based upon volunteer training (including coaching), some of which is to be provided by local authorities with the help of the Sports Council.

These initiatives are replacing the traditional local authority unconditional support of sports clubs by conditional support. Under new schemes, local and national resources are only made available to sports clubs in return for the sports club adopting strategies that contribute to the strategic objectives of both government and the local authority. Support is now conditional upon the sports club adopting business practises not only

\textit{SPORTS COACHING}
to improve effectiveness, but also to allow auditing of the attainment of the strategies, so called “accountability”. The business practices being required include published plans; formal management structures; written job descriptions; formal procedures for recruitment and training of volunteers; and marketing, publicity and fund raising posts.

This changing role of local authorities from unconditional to conditional provision of support is leading to tensions. Even the authors of the national survey that stimulated this change two years later warned that the initiatives may be threatening the pluralistic diversity of voluntary sports clubs (Nichols, G, and others, 1998). Voluntary coaches are beginning to react against this bureaucracy. The Motion that “This House believes the Spirit of Sport is dying” was debated in the Oxford Union with David Train, an Olympic canoeing coach, supporting the Motion and Sir Christopher Chataway opposing the Motion. The Motion was carried (Train, D, 1998 (a) and (b)).

It can be argued that business models are not appropriate for many voluntary bodies. Business models are predicated upon clear, relatively stable and financially measurable aims and the ability to recruit skills as and when required. Voluntary bodies respond to the dynamics of perceived ‘gaps’ in society and are geared to utilising and working within the local skills and motivations presented to them.

Differing views have been expressed on the future of the volunteer or amateur in sport. In 1980 a European study specifically called for the UK voluntary tradition to be respected and maintained, but advocated increasing professional involvement (Anthony, D, 1980). The study warned about the dangers of power motivation of amateur officials. It also suggested that the voluntary clubs would become adjuncts of companies, commercial clubs and of schools.

This has been incorporated in the English Sports Council strategy (English Sports Council, 1997). This advocates partnerships between voluntary clubs and the public

SPORTS COACHING
and private sectors and improving the effectiveness of volunteers by positive incentives to take formal training in business practices. Encouragement and assistance to train are planned to come from the employment of sports development officers by the NGB’s and by local authorities. Further incentives have been introduced through the National Lottery. Applications now require business plans showing how partnerships are to be forged and how clubs will run. Whilst there is acknowledgement that sports differ from each other, the partnership and organisational models for clubs are standard across all sports.

Psychological research suggests that there are some common characteristics in participants across all sports (Butt, D S, 1987). These characteristics do not vary much with gender or between amateurs (volunteers) and professionals or, possibly, between competitors and officials. All exhibit a degree of egocentricity and a lack of consideration for others. Each sport can be placed on a two dimensional plane made up of the degree of extroversion (or, negatively, introversion) and of degree of stability.

However, other research (Lyle, J, Allison, M, and Taylor, J, 1997) has found that specific sports have their own patterns. Group or team sports tend to have participants who are more extroverted, have more self confidence, are more motivated by demonstrating their skill competence and are more able to handle defeat. Individual sports attract people who are less extroverted, less confident, require external support and have difficulty in handling defeat. There is some evidence that extroversion and stability increase in an individual with length of their participation. Whilst women in sport are less extroverted than males in sport, they are still more extroverted than women not involved with sport.

Each sport has developed its own culture. Within each culture, a set of sub-cultures usually arises. Each sub-culture had its own set of rules, code of conduct, dress, hierarchy and even physical space. Evidence for this model can be seen at any football match (players, officials, managers, Directors, stewards, etc.) or any athletics
meeting (athletes, starter, time keepers, officials, coaches, etc.). Members of each sub-culture see themselves as central to their sport. They avoid conflict with the other sub-cultures by not trespassing on their space. The sub-cultures share in common appreciation of skill and of excellence. By these means, each sub-culture satisfies the needs of its members (Butt, D S, 1987).

The theory indicates that any research trying to identify motives and deterrents to participation must make distinction between:

- the type of participation, coaching being one type;
- sports, at least between team and individual based sports;
- gender;
- length of participation in sport.

A study in the USA (Wheeler, A J, 1998.) validated a multi-dimensional scale for measuring the motivation of a coach. Wheeler found that a coach’s Task orientation related to:

- Belief that the purposes of sport are to teach athletes the importance of adopting a healthy lifestyle
- To teach athletes to be good citizens
- To enhance athletes self-esteem
- To teach athletes how to be co-operative and master skills
- To teach athletes to be competitive.

Ego orientations were associated with:

- Belief that the purposes of sport are to teach athletes to be competitive
- To enhance the social status of athletes
- To prepare athletes for high-status careers.

Task orientations were found to be related to 4 factors (enjoyment, satisfaction, interest and commitment), but ego orientations were found to be not related to these factors.
Wheeler's research confirms and reinforces the fact that being a sports coach is a multi-dimensional process, the research findings may not be applicable to voluntary coaching as the data related to professional coaches, mainly at academic institutes. The motives of such professional coaches are likely to reflect their personal desires for career progression and establishing their personal identities with the most promising students and with the academic institute employing them. These desires can be seen reflected in the majority of the factors listed above. They indicate that the scale may relate more to a category of professional employment than it does to sport as an activity.

It should also be noted that there is an emphasis on collaboration, i.e. a team environment. This, too, is a characteristic of US academic sports teams headed by a professional coach. The task of the coach, through the team, is to enhance the reputation of the institute by the collective performances of the team. The athletes comprising the team may be engaged in a team sport, such as baseball, or an individual sport, such as athletics. The characteristics of the sport are of lesser importance than the results as a team.

With the logistic regression modelling framework presented in this research, the focus is entirely upon voluntary coaching and with one sport, athletics. This is a sport based upon individual achievement. However, competition, especially at club level, is organised around teams. The teams compete in leagues with relegation and promotion rules between divisions within each league. Clubs frequently identify themselves as much with their league positions as with the individual successes of their athletes. The athletics coach in a club setting has to be concerned both with the individual characteristic of the sport and the fact that the sport and the coach are operating in the club environment that encourages a team approach. This contrasts strongly with the professional coach employed for team results in the US academic setting reported in the research by Wheeler.
The amateur athletic coach may come with a history and experience of personal participation in an individual sport or of participating as a member of a team. This history is likely to effect the individual perception of and approach adopted to coaching. The degrees to which these team aspirations exist are likely to vary from club to club.

Psychological research has also examined what happens to players when they cease competing. It suggests that players tend to maintain contact with their sport to preserve their identity. They may coach or they may introduce their children or just keep in touch. A minority may decide to pursue another sport in order to find pleasure through other skills (Butt, D S, 1987). This concept is significant if a high proportion of volunteers in any sport are people who have previously participated in that sport. If so, that sport will maintain the characteristics of that set of people. Recent developments of Masters and Veteran sports now allow competition to continue throughout life. This may result in adults choosing to remain participating as Masters or Veterans to remain within the sport, rather than moving to coaching, officiating or other administrative activities as many did in the past.

3.4. Sport, Government and Coaching

Histories of sport in the UK (Holt, R, 1989; Houlihan, B, 1991; Cox, R, 1994) show that, until the 19th century, it was largely an amateur, non-commercial, leisure time activity organised and played on a local basis by volunteers, usually participants, as a respite from and contrast to normal work, a means of personal re-creation of the spirit. From time to time, attempts had been made to use sport for social purposes:

- the state to ensure an adequate supply of skilled archers;
- the church to ensure observance of the sabbath;
- education to instil discipline and leadership.

But all of these attempts either fell into disuse or were exercised over targeted sections of the population.
Improvements in transport and communication in the 19th century changed the local nature of sport into national and, by the end of the century with the birth of the Olympic movement, international domains. These required each sport to be codified and for national and international governing bodies to be set up. In some sports, professionals became involved either to enhance the success of local teams or to teach the sport. The national introduction of formal rules and organisation began to be reflected in the greater formality of organisation at local level. As sport was still essentially non-commercial, the local organisation was undertaken by unpaid volunteers. In most sports, it was these same volunteers that elected and formed the workforce of the national organisations and determined policy.

In the first half of the 20th century, sport began to take on a national significance, especially with the growth of the Olympic movement. This led governments to take an interest, an example being the 1936 Berlin Olympics. Increases in leisure time and disposal incomes, emancipation of women and increases in school leaving ages increased the amount of participation. It also led to an increase in the number of professionals involved, both as administrators and coaches. The growth and influence of sport began to attract the attention of commerce, particularly the media and the clothing and footwear manufacturers. But the control of each sport largely remained with elected volunteers.

Both world wars led to increased interest in physical education for the training of the armed services. At the same time, science was shedding increasing light on how people moved, learnt and improved their fitness. These enabled a profession of physical education to be borne. Their influence was mainly channelled through schools that were seen as the right place for children to be trained to be healthy adults and to be introduced to sport. Physical educators had little direct influence in local sports clubs and teams. Physical educators were largely trained in colleges of education for teaching in primary and secondary education.
During the second half of the 20th century, governments encouraged national governing bodies of sport to improve their successes at international level. This led to the increased employment of professional administrators and coaches. In some sports, for example football, this filtered down to local teams. In others, for example athletics, it remained at national level only. The growing employment of professional coaches led to demands for formal training for coaches. During the same period, there was a rise in the number and sizes of Institutes of Higher and Further Education. Some of them, for example Birmingham and Loughborough, looked for new disciplines to establish an identity and to attract students. Sports science and sports sociology became accepted academic disciplines. Nevertheless, the majority of administration and coaching in the UK is still undertaken by volunteers (The Sports Council, 1996(a)). The main initial motivation for a volunteer to coach appears to be to assist their own children or their own club to compete more effectively.

During the 1960’s to 1980’s, encouraged by their professional coaches, national governing bodies began to design and implement formal education programmes for amateur coaches at local level. The course contents were largely written by the academically trained professional coaches. This process was endorsed by the formation by the Sports Council of the National Coaching Foundation (NCF). This body has established a nation wide programme of non sport specific education modules aimed both at coaching and sports administration. Where a NGB has its own sport specific coach education programme, the NCF modules are complimentary, covering non sport specific areas such as child development, injuries, drugs, legal issues, finance, etc. Over the past few years, an increasing number of modules have been aimed at communicating new Sports Council concerns, including Valuing Volunteers and Child Abuse.

During the 1990’s, pressures increased for coaches to be qualified. Insurance companies made qualification a requirement for the club being covered. Local authorities and local education authorities made qualification a stipulation for being involved in coaching in local authority schemes and at schools. Local authorities
made the presence of an appropriately qualified coach a condition for use of their facilities.

Prior to the 1990’s, coaches could operate without any prior screening or reference. Increasingly local authorities and NGB’s are requiring all potential coaches to be screened in accordance with the Child Protection Act. Positive screening then becomes a prior requirement for entering a coach education programme.

It will be seen from this brief history of coaching that both the actual and perceived roles of a coach have changed over time. The previous Chapter (2) showed how the roles of a volunteer had also changed. Any decision, including that of becoming a coach, is influenced by the perceptions as well as facts surrounding the individual at the time of decision making. In broad terms, perceptions are primarily related to the age of the individual, their historical generation, although these will be altered by their life experiences. Facts, on the other hand, are broadly those that the individual has access to around the time of decision, although these will also be influenced by life experiences to that time.

Ideally, at least four variables should be used to cater for these two effects: generation, background, historic time of decision and life experience. Each is a complex variable. Only proxies are, realistically, available. For example: date of birth, a socio-economic metric or metrics, historic date of decision and an education/work metric or metrics. In practise, any set of such proxies is likely to present statistical problems as they are likely to be correlated amongst themselves, real dates and history being common factors. The theoretical problems associated with correlation extend further as other variables may also be correlated. An example in this research are age (date of birth) and whether the individual has dependants.

Whist the research will measure proxies for the four variables discussed in the previous paragraph, the degree of correlation between variables found in the analysis will determine whether or not a specific variable is used. This is discussed in detail in 

SPORTS COACHING
Chapters 6 and 7. The result is that one or more important variables may be omitted from the final model. This will be take into consideration in the interpretation of the research results.

The next Chapter looks at the UK Athletics Coach Education Programme. This will give some indication of how volunteers may perceive the programme and, hence, the motivations and deterrents that the Programme may present to them.
4. THE UK ATHLETICS COACH EDUCATION PROGRAMME

4.1. Introduction

The Athletics Coach Education Programme in the UK was introduced nationally in the 1970’s. A major review took place in 1999 and significant changes were introduced from the year 2000 (UK Athletics, 2000). The data collection for this research took place in 1999 before these major changes. The Programme described in this Chapter refers to the Programme prior to 2000.

UK Athletics (UKA) is the NGB of athletics in the UK. It was formed in 1998 to replace the British Athletics Federation (BAF) which, in turn, had replaced the British Amateur Athletics Board (BAAB). The athletics coach education programme is specific to the needs of athletics in terms of the various disciplines (running, throwing, jumping, hurdling, walking, etc.).

The Athletics Coach Education Programme in the UK is separate from the National Coaching Foundation (NCF) programmes which are not sport specific and cover the administration of sport as well as coaching. NCF administrative areas include young children, child protection, legal & insurance, recruitment, etc., many of which impact on the role of coaching. Whilst attendance on NCF courses is recommended, in practise 2-3% of athletics coaches attend them (Lyle, J, Allison, M., and Taylor, J, 1997). Surveys of Level I courses Crystal Palace 30/11/97 & Horsham 15/2/98). For this reason, the NCF programme has not been considered in this research.

Whilst the content, format and details of delivery have been revised a number of times, the overall structure and subjects covered had not been change significantly (until 2000). Alterations reflected technical and medical progress and presentation improvements. Between 1995 and 1999, the length of courses increased and some qualifications became valid for specified periods of time. The structure, detailed
course contents, tutor instructions and printed course manuals were decided and produced at national level. The delivery was planned and implemented at regional level. Consequently, there were some minor differences in delivery. The delivery was by formal lectures and practical courses. Courses were held at numerous locations throughout the country, usually within 50 miles of any local club. Each course usually occupied one or two weekends. Participants had to provide written work and to demonstrate their practical competence. There was a written examination, varying from a multiple choice question sheet at the entry level to 2 hour written papers at senior levels. The researcher was led to understand by UKA that the failure rate (excluding drop out) was less than 2%. The course tutors were mainly experienced, voluntary coaches under the guidance of the regional, professional coach. Tutors were provided with notes, but most tutors were not formally trained to deliver the Programme. The course contents and the manuals were written by professional coaches and physical educationalists. Fees were payable and included the relevant manual. Participants were expected to consult other text and reference books, either by borrowing them or by purchase.

4.2. Recruitment

Communication within athletics in the UK is critically dependent upon the local club. There is no national newsletter. Information is circulated from UKA to the secretaries of individual clubs with some information emanating at regional level. A limited amount of information is circulated directly to qualified coaches. Until 2000, there was no central database of club members or of qualified coaches. Some information, including coaching courses is advertised in the two national weekly magazines devoted to athletics and running. However, the readership of these two magazines is probably less than 10% of the population associated with athletics.

Clubs can obtain leaflets and posters concerning coach education. There are also packs enabling a club to run introductory evenings for members and parents. Clubs are also encouraged to give new members an introductory sheet, including facts about
coaching. What advertising is undertaken is often swamped by other information on a notice board.

In practise, information about the Coach Education Programme largely reaches members and parents by word of mouth from club officials and existing coaches. As a result, most individuals only receive information about courses if they have shown or expressed some interest in becoming involved with the club as a volunteer or because they express concern about coaching for themselves or their children.

The main exception to this situation is with teachers and students with an interest in a career in physical education or in the sport and leisure industry. They may seek for accepted qualifications to aid their work or to improve their CV’s. These people are more likely to be proactive and seek information either from an athletics club or from the NGB.

4.3. The Athletics Coach Education Programme

From the Autumn of 2000, UK Athletics significantly altered both the structure and the contents of the coach education programme. This Section concerns the coach education programme available before the year 2000. The main programme consisted of three levels:

- Level 1 (originally called Assistant Club Coach),
- Level 2 (originally called Club Coach),
- Senior Coach.

All coaches started at Level 1. Thereafter, they could not upgrade until they have completed the previous level and have a valid qualification in that level. From 1997, qualifications were only valid for a set period. A Level 1 qualification expired after 6 years and has to be re-validated every 2 years. A Level 2 qualification expired after 4 years and had to be re-validated after that.
As this research only concerns the initial decision to enter the Coach Education Programme, only the details of the Level 1 awards are considered further in this Chapter.

The decision to take a coaching award required a choice of events to be studied. There were broad event groups. The choice for Level 1 was between these groups. The groupings that existed in 1999 for Level 1 are shown in Table 4.3.1.

Table 4.3.1. 
ATHLETIC EVENT GROUPS FOR LEVEL 1

<table>
<thead>
<tr>
<th>EVENT GROUP</th>
<th>EVENT SUB-GROUPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPEED</td>
<td>Running up to 400M &amp; Relays</td>
</tr>
<tr>
<td></td>
<td>Hurdling up to 400M</td>
</tr>
<tr>
<td>ENDURANCE</td>
<td>Running: 800M &amp; 1,500M</td>
</tr>
<tr>
<td></td>
<td>Running: 3,000M to 10,000M</td>
</tr>
<tr>
<td></td>
<td>Marathon Running</td>
</tr>
<tr>
<td></td>
<td>Race Walking</td>
</tr>
<tr>
<td></td>
<td>Fell Running</td>
</tr>
<tr>
<td></td>
<td>Steeplechase</td>
</tr>
<tr>
<td>JUMPS</td>
<td>Long &amp; Triple Jump</td>
</tr>
<tr>
<td></td>
<td>High Jump</td>
</tr>
<tr>
<td></td>
<td>Pole Vault</td>
</tr>
<tr>
<td>THROWS</td>
<td>Shot Putt/Discus</td>
</tr>
<tr>
<td></td>
<td>Javelin</td>
</tr>
<tr>
<td></td>
<td>Hammer</td>
</tr>
<tr>
<td>COMBINED</td>
<td>Heptathlon/Decathlon</td>
</tr>
</tbody>
</table>

This specialisation only effected the practical elements of the course. The written manuals concern theory and are non event specific. This minimised any effects in the study of the event chosen by each individual, providing that random sampling is used to confound any effects of personal characteristics and chosen event.

There were conditions imposed for entry to each level of coaching award. The conditions for Level 1 are shown in Table 4.3.2.
Table 4.3.2.
CONDITIONS FOR AWARD - LEVEL 1

<table>
<thead>
<tr>
<th>CONDITION</th>
<th>LEVEL 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MINIMUM AGE</td>
<td>18</td>
</tr>
<tr>
<td>PRE-REQUISITE</td>
<td>None</td>
</tr>
<tr>
<td>PRIOR LEVEL</td>
<td>None</td>
</tr>
<tr>
<td>COURSES</td>
<td>General Coaching</td>
</tr>
<tr>
<td></td>
<td>All Event Overview</td>
</tr>
<tr>
<td></td>
<td>Event Group</td>
</tr>
<tr>
<td>Number of Days</td>
<td>3</td>
</tr>
<tr>
<td>WRITTEN</td>
<td>Multiple choice</td>
</tr>
<tr>
<td>EXAMINATION</td>
<td>None</td>
</tr>
<tr>
<td>ASSIGNMENTS</td>
<td>Reading</td>
</tr>
<tr>
<td>ASSESSMENTS</td>
<td>None</td>
</tr>
<tr>
<td>COACHING</td>
<td>12 Sessions</td>
</tr>
<tr>
<td>COMPLETION</td>
<td>6 months</td>
</tr>
<tr>
<td>REVALIDATION</td>
<td>Every 2 years</td>
</tr>
<tr>
<td>EXPIRY</td>
<td>Maximum 6 years</td>
</tr>
<tr>
<td>FEE</td>
<td>£55</td>
</tr>
<tr>
<td>(South of England Region, 1998)</td>
<td></td>
</tr>
</tbody>
</table>

In relation to this research, the following are important:

- Reading. Level 1 candidates have to sign a declaration that they have read the manual.

- Only adults, individuals aged 18 or over, can qualify as coaches

- The level of fee, although some clubs will pay some or all the fee, possibly dependent upon the individual obtaining the qualification and providing the club with a minimum amount of coaching.

- The amount of time both to take the course and to do the necessary amount of reading. The latter is further considered below.

A Level 1 Theory Manual was issued at the time of registering for the course and the cost was included in the fee. The contents of this manual are summarised in Table 4.3.3. It will be seen that about half of the contents are devoted to topics with a technical basis, e.g. strength conditioning, in which scientific terms may occur.
Table 4.3.3.
THEORY MANUAL - LEVEL 1

<table>
<thead>
<tr>
<th>TITLE</th>
<th>PAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>So now you wish to coach</td>
<td>2</td>
</tr>
<tr>
<td>Guidelines for coaching conduct</td>
<td>4</td>
</tr>
<tr>
<td>The organisation of British athletics</td>
<td>6</td>
</tr>
<tr>
<td>Athletics for all</td>
<td>7</td>
</tr>
<tr>
<td>Coaching athletes with disabilities</td>
<td>4</td>
</tr>
<tr>
<td>Body structure</td>
<td>5</td>
</tr>
<tr>
<td>Childhood to veteran</td>
<td>6</td>
</tr>
<tr>
<td>Athletics as a part of the athlete’s life</td>
<td>5</td>
</tr>
<tr>
<td>Athletic games</td>
<td>6</td>
</tr>
<tr>
<td>Safety in athletics</td>
<td>5</td>
</tr>
<tr>
<td>Mobility training</td>
<td>4</td>
</tr>
<tr>
<td>Technical training</td>
<td>7</td>
</tr>
<tr>
<td>Endurance training</td>
<td>4</td>
</tr>
<tr>
<td>Strength conditioning</td>
<td>11</td>
</tr>
<tr>
<td>Speed training</td>
<td>4</td>
</tr>
<tr>
<td>Planning training</td>
<td>6</td>
</tr>
<tr>
<td>Sources of information and reference</td>
<td>2</td>
</tr>
<tr>
<td>TOTAL PAGES</td>
<td>88</td>
</tr>
<tr>
<td>TOTAL WORDS</td>
<td>29,000</td>
</tr>
</tbody>
</table>

(Warden, P, 1996)

Most Chapters of the manual have further reading lists. The individual student was expected to obtain further reading through libraries or by purchase.

Clearly, the Level 1 course required a student to undertake a substantial amount of reading in a wide variety of subjects, some of them technically and scientifically orientated. There was an assumption that the level of literacy required could be expected from potential volunteers. Overall reading ability or, more specifically, ability to read technical and scientific articles could be a potential variable in the decision process. How this was incorporated in the questionnaire used in the interviews is discussed in Chapter 8.
5. DECISION MODELS

5.1. Social Exchange Theory

Chapter 2 reviewed the relevant research regarding volunteering and introduced the concept of Expectancy Theory in Section 2.6.

A development of Expectancy Theory that makes it more relevant to sports clubs is known as Social Exchange Theory. The development is the incorporation of a dynamic element into the choice process. In Social Exchange Theory, incentives are based upon exchange of costs and rewards within a continual process of evaluation of expected and actual costs and benefits. The process can be seen as having five stages:

i. generation of ideas leading to expectations, especially from many forms of publicity;
ii. initial decisions and explorations, resulting from the raised expectancy;
iii. specific decisions following perceptions of benefits and costs as a result of explorations;
iv. acting on these specific decisions and also continual balancing of actual benefits and costs against expectations and perceptions;
v. repeated decisions resulting from the balancing on whether or not to continue acting or to act again.

(Phillips, M, 1982).

This process and its five stages are represented, in general terms, in Diagram 5.1.1.
The complete process, from idea to specific action, takes place over time and is dynamic. In the process, ideas lead to expectations that, in turn, generate initial decisions. These decisions are then explored for benefits and costs. Specific decisions follow, including doing nothing, and specific actions taken to implement the decisions. At the same time, a review process is started of balancing the perceived benefits and costs against the perceived expectations. If either set of perceptions differs from those anticipated or are perceived to have changed, then the process is
restarted. Differences may arise from changes in individual circumstances or the environment. A new review may be instituted whenever there is an actual or a perceived change in the benefits, costs or expectations. Changes may be direct from the environment. Changes can also be indirect as a result of alterations of the perceptual scales of the individual from new experiences.

Social Exchange Theory and its five stages can be put into the specific context of voluntary sport and coaching by considering an individual or the parent of an child.

Stage 1. The person has an idea about sport leading to expectations of either fitness (including competition) or social gain through sport (or both).

Stage 2. The initial decision is taken to join a club with a view to gaining well-being or improving social status (or both).

Stage 3: An exploration is made of the club and perceptions of benefits and costs are formed. These lead to considerations of personal improvements in sporting performance or social status (or both).

Stage 4: A parent may decide that improvements for their child, possibly themselves, can be expected by becoming a coach. An adult may decide that personal improvements can be achieved by becoming a coach.

Stage 5: The benefits, costs and expectations of becoming a coach are continuously reviewed.

This process is shown in Diagram 5.1.2. and can be compared to Diagram 5.1.1.

DECISION MODELS
Diagram 5.1.2.
SOCIAL EXCHANGE THEORY IN VOLUNTARY SPORT

The decision process is dynamic in that it is continuously being effected by the environment made up of and impacting the individual and the sport.

The research in volunteering and sport which was reviewed and cited in Chapters 2 and 3 have found significant variables that can be grouped into 8 areas:

1. Personal variables: gender, age and sporting experience.
ii. Educational factors: school, further and higher education, training and literacy.

iii. Family factors: composition and the sporting interests of individuals, especially children.

iv. Work factors: regularity, work time and commuting.


vi. Publicity: forming attitudes towards volunteering and coaching.


viii. Heath: sports involvement.

These 8 areas surround, pervade and effect the components of the decision process. Not only is this a dynamic, multi-dimensional environment, there is considerable interaction between the 8 areas. The Social Exchange Theory applied to Sport in Diagram 5.1.2. can then be put into context surrounded by this environment, as shown in Diagram 5.1.3.

This research is only concerned with that part of the model relating to the decision to become a coach. In terms of the model conceptualised in Diagram 5.1.3. the research is concerned with the box labelled “Improvement” and the specific outcome action of “Be a Coach”. However, the 8 environmental sets effect this decision because they effect costs and benefits leading directly to the decision. They indirectly effect this decision through all the other parts of the process from the initial generation of the idea about sport at the top of Diagram 5.1.3.

The box labelled “Balance” relates to the subsequent decisions of progressing to Level 2 and beyond or staying at Level 1 or to the decision to cease coaching. Whilst these processes of progression, retention and termination are of great interest as they determine the number and quality of the coaching pool, they are not the immediate concern of this research. However, some of the qualitative data gained by the research does shed some insight into these and is reported in Chapter 12.
Diagram 5.1.3.
SOCIAL EXCHANGE THEORY FOR COACHING

PERSONAL  EDUCATION  FAMILY

SPORT

FITNESS or SOCIAL

JOIN CLUB

PROGRESS or STATUS

BENEFITS & COSTS

IMPROVEMENT

BECOME A COACH

BALANCE

FINANCE  PUBLICITY  SOCIETY
The conceptual model in Diagram 5.1.3. lacks a specific mechanism for determining whether or not the decision by an individual leads to specific action to enter the coach education programme. Without such a mechanism, the individual may have made a decision, but not put it into action. This mechanism is introduced by incorporating a Decision Threshold into the model. This is now discussed in Section 5.2.

5.2. The Decision Threshold

A conceptual model of how people adopt recreational activities was developed by Brandenberg and others (1982). They set out to determine how the preoccupations and interests of an individual became linked to a specific activity. From the beginning, they made a distinction between decisions to start participating in a specific activity and decisions to remain in a specific activity or to cease participation. This distinction is analogous to the distinction between the decision to become a coach, the interest of this research, and later decisions to advance in level of coaching or to cease coaching, neither of which are the concerns of this research.

Their definitions of ‘preoccupations’, ‘interests’ and ‘activities’ followed those in the earlier concept of Family Life Cycles (Rapoport, R, and Rapoport, R N, 1975). Brandenburgh and his associates found evidence that these preoccupations and interests stemmed from various conditions that could be grouped into four elements: Opportunities arising from favourable social and environmental factors or absence of constraints that included accessibility, transport availability, physical capability, financial feasibility, time availability, necessary resources and changes in personal circumstances.

Knowledge of the activity, or at least a vague idea, which generated, at least, some curiosity. This knowledge might come from prior experience of the activity, from reading or being told about the activity, or seeing it live or on TV. Perceiving the activity to relate favourably to or, at least, acceptable to the individual’s reference group such as family or peers. The new activity must also be
perceived as being socially compatible with the individual’s existing interests, activities and values. Brandenburg et al. referred to this as ‘the social milieu’.

Willingness or desire to participate in the activity, referred to as ‘receptiveness’.

Their model hypothesised that the decision to take up an activity was triggered by the occurrence of what they termed a ‘key event’. These ‘key events’ could also effect one or more of the four elements listed above. A ‘key event’ might also determine the specific activity to be entered. They suggested that the occurrence of a ‘key event’ could cause the individual to reassess the four elements. The model was tested using a phenomenological approach through a series of interviews and triangulated from the results of a written questionnaire. They found that every participant reported at least one ‘key event’ that had triggered them adopting a new activity. The ‘key events’ included many of the events identified by Rapoport and Rapoport in their study of Family Life Cycles, viz. Change of job, move of residence, marriage, birth, retirement. In addition, they found that ‘key events’ could result from being personally asked to participate, being given a piece of equipment, seeing striking publicity, accidentally attending an activity, etc. A ‘key event’ could also arise from a remark or incident unconnected with the activity making the individual realise that they were bored with their present set of activities.

The Brandenburg model is shown in Diagram 5.2.1. This bears some relationship to the research model presented in Section 5.1. and illustrated in Diagram 5.1.3. There is strong resemblance between the occurrence of a ‘key event’, discussed in Section 5.2., and the generation of the idea at the beginning of the model presented in this research. The four elements of conditions for interest are the equivalent of the 8 areas of environmental factors in the research model around the outside of Diagram 5.1.3.
However, the Brandenburg model relates to the whole process described in that Section, but not specifically to the mechanism underlying the actual decision outcome.

This deficiency of the Brandenburg model was addressed in 1994 in the work of Mintern (1994) looking at non-elite individuals deciding to use and to cease using a new leisure centre in the UK. He took the opening of the leisure centre to be a ‘key event’. His research centred on modelling how individuals reacted to this ‘key event’ by deciding or not to participate in the leisure centre activities and then whether to remain or drop out. The research specifically did not concern itself with why they made decisions. The research took the Brandenburg model and then hypothesised that the variables effecting the decisions were additive in their effect (positive if motivating, negative if constraining); that each variable was specific to the individual; and that each variable had its own degree of importance. His theory was that there was a ‘threshold’ value for the total of the effects above which the decided to participate and below which the individual decided not to participate.
Mintern went on to group the variables effecting the situation into four levels:

- **Level 1** - The individual's basic self perception, abilities and intrinsic motivators
- **Level 2** - The individual's social network, lifestyle, resources and finances.
- **Level 3** - The degree of accessibility of the leisure activity to the individual in terms of time, geographic location and availability
- **Level 4** - The novelty of the leisure activity to the individual.

In addition, he included a cost variable that was expected to act as a constraint.

The basic Mintern model is conceptualised in Diagram 5.2.2. showing the motivation and threshold of the individual prior to, at the time of and subsequent to the opening of the new facility. Prior to the opening, the individual decides not participate in the facility. The individual decides to join at the opening. Later, the decision is reversed and the individual decides to drop out. These decisions are conditioned by relation of the motivational value and the threshold. The motivational value and the threshold vary with the changes over time of the 4 Level variables, the cost variable and the threshold.

The left hand side of Diagram 5.2.2. shows the situation for the individual prior to the opening of the new leisure facility. The individual’s pre-opening motivation level (shown as a dashed horizontal line) to participate in recreational activity is determined by the additive effects of the Level 1 to Level 3 variables. The individual has a threshold value relative to the motivation level. The individual will only decide to take part in the activity if the motivational level exceeds this threshold. The value of the threshold is personal to the individual, but is also effected by the situation, especially by the cost variable. The cost variable acts as a constraint to the individual and raises the threshold value. In Diagram 5.2.2., prior to the opening of the new resource, the individual does not participate as the pre-opening motivation level does not exceed the threshold.
Diagram 5.2.2.
MODIFIED BRANDENBURG MODEL - THE SPORT MODEL

(after Mintern, J.H., 1994, p. 284)
The left hand vertical dashed line indicates the opening of the facility. To the right of this the Level 3 variable (accessibility) increases and the Level 4 variable (novelty) is introduced. The net result is that the level of motivation rises to its post-opening level. The cost variable increases as the new facility is more expensive. This increase in the cost variable raises the threshold to its post-opening level. However, on the opening of the new resource, the level of motivation has increased by the better accessibility and by the novelty and exceeds this new, post-opening threshold, despite this also being raised by the high costs. The individual decides to participate.

As time progresses, any or all of the Level 1 to Level 4 variables and the cost variable may change. Any or all of the changes to the Level 1 to Level 4 variables may raise (or lower) the motivational level of the individual and the cost variable may raise or lower the threshold. To the right of the vertical line indicating the facility opening, the Level 3 (accessibility) and Level 4 (novelty) variables are shown decreasing. As a result, motivational level decreases. Neither the cost variable nor the threshold change. At the point of the right hand vertical dashed line, the motivational level falls below the threshold. This may result in the individual then deciding to drop out of the activity.

The Mintern model bears many analogies to the decision situation in deciding to become a voluntary athletics coach. It was also derived from the Social Exchange Theory. The next Section, 5.3., shows how the Mintern model has been adopted to the process considered in this research of the decision to become a voluntary athletics coach. This then becomes the model on which this research is based and whose relevance will be demonstrated by the degree of fit achieved when it is tested.

5.3. The Sports Model

A Sports model was derived by taking the 8 factors surrounding the Social Exchange model for Coaching shown in Diagram 5.1.3. and placing them in the framework of DECISION MODELS
the modified Brandenburgh model shown in Diagram 5.2.2 as replacements for the 4 level variables of Mintern. Mintern’s cost variable becomes part of the variable ‘Finance’.

The motivation Threshold will be exceeded when the net sum of the perceived benefits of all the motivating variables exceeds the value of the threshold. Each Threshold is associated with ‘Key Events’. The decision is taken when a ‘Key Event’ occurs after the motivation Threshold has been exceeded.

An example of this Sports model is shown in Diagram 5.3.1. The example is deliberately simplistic in order to illustrate a complex and multivariate process. In this example, the individual is assumed to be a parent. For the sake of clarity, 4 of the 8 factors are assumed to remain constant over the time span of Diagram 5.3.1.: Personal, Education, Work and Society. In reality, changes over time are likely to all these variables. Family is assumed to increase its motivation over time, as a result in the individual wishing to enhance their child’s health and social experience. Publicity is assumed to increase motivation, as contact with the sports club increases over time. Health is assumed to increase with time, as awareness of health issues increases with age and with media attention. Finance is assumed to increase motivation, as the resources available to take part in the sport increase with time.

Two thresholds and Key Events are shown in Diagram 5.3.1.:  
1: triggering a decision for the child to join a sports club  
2: triggering a decision for the individual to become a coach.

Increases in the variable ‘Family’ occur as the parent places emphasis on the child’s health and socialisation. The variable ‘Publicity’ increases as knowledge of the club comes from the child’s school, local newspaper reporting and friends talking. Motivation increases to exceed Threshold 1. The Key Event might be the child’s PE teacher suggesting that the child should join the athletics club. The parents enrol the child in the club. These occurrences are hypothetical and used illustrate the process.
Diagram 5.3.1.

SPORTS MODEL ADAPTED FOR COACHING

Key Event 1
Join

Threshold 1

TIME

Key Event 2
Coach

Threshold 2

MOTIVATION

HEALTH

FINANCE

PUBLICITY

SOCIETY

WORK

FAMILY

EDUCATION

PERSONAL

A variation of this model would be that it succeeds, only an individual in such an adaptation of the Sports model. Threshold 1 would be reached by the individual becoming a coach to become aware of the health issues through coaching the sport. The official of the club asks the parent to help in coaching. The parent then decides to become a coach and decides to become aware of the health issues through coaching the sport.
The variable ‘Family’ continues to increase as the child gets pleasure from and attains success in the club. ‘Publicity’ increases as the child brings home news from the club and medals from events. The variable ‘Health’ begins to increase as the parent becomes more aware of health issues through doctors and the media. The combination of these increases leads to Motivation exceeding Threshold 2. Latter, an official of the club asks the parent to help in coaching. The parent then decides to become a coach to further the improvements of the child at the club.

A variation of this model would be that it concerns only an individual. In such an adaptation of the Sports model, Threshold 1 would be associated with the individual joining the club for their own sports participation. Threshold 2 would continue to be associated with that individual deciding to become a coach.

Both these adaptations of the Sports model (parent and child; individual) have been used in this research. An assumption is made Threshold 2, the decision to become a coach, has a constant value across both adaptations. The results of the analysis of the model from data collected from parents and from individuals will indicate the extent to which this assumption is acceptable.

Threshold 1, the decision to join a club, shown in the example of the Sports model will not be considered in this research. The research concentrates on one decision, to coach or not coach. It also concentrates upon the attainment of the threshold.

The same model could be extended in time to decisions regarding the parent upgrading the coaching qualification as ‘Motivation’ rises as a result of greater participation as a coach. Further along time, some variables may begin to decline. For example, the child may cease participation at the club as a result of examinations or leaving home for work or higher education. This may lead to the variable ‘Family’ decreasing and, possibly, the variable ‘Finance’ decreasing as a result of having to fund further education of the child. The health of the parent might deteriorate decreasing the value of the variable ‘Health’. As a consequence, another Threshold

DECISION MODELS
associated with coaching might be reached and the parent takes the decision to cease coaching.

The model can be made applicable to many sports and to many activities in a sport, including the individual’s competition career, serving on a club committee and acting as an official.

The Sports model described in this Section does not attempt to address three other components:

i. How the value of a Threshold specific to a decision is determined in relation to Motivation

ii. The relationship of Thresholds with ‘key’ events

iii. The specific ‘key’ events associated with each Threshold.

The reasons for the omission of these three components are twofold:

- the research is restricted to one specific Threshold. The methodology of logistic regression used in this research to solve the model for a specific population allows thresholds to be set, tested and, if necessary, adjusted.

- the research is only concerned with the associations of variables to making of decisions on coaching and not why decisions are made. The research is concerned with the pre-requisites for making a decision. ‘Key’ events are related to when individuals make decisions in time and why they make them.

The Sports model as illustrated in Diagram 5.3.1. is simplified in two respects.

- It indicates that motivation does not decline below the threshold between it initially exceeding the threshold and a Key Event occurring.

- It also assumes that a Key Event does occur.

If motivation declines below the threshold before a Key Event occurs, then it would be expected that the decision is not made. Similarly, if a Key Event does not occur,
then also it would be expected that the decision is not made. These illustrate the importance of Key Events and their timing in relation to motivational level.

Having derived a Sports model and adapted it to a specific situation, deciding to be an athletics coach, the next Chapter describes the methodology for solving the model for a given population.
6. STATISTICAL METHODOLOGY

6.1. Introduction

The limited surveys of sports coaching in the UK have used forms of comparative methodology to compare characteristics between groups selected from surveys. Many of the comparisons have been on one variable only. Limited comparisons have been made with groups of variables selected from cross tabulations of two variables (Lyle, J, and others, 1997; The Sports Council, 1996(a)). The same types of methodology have been used in the major UK research programmes to determine the characteristics of volunteers both in sport and other social activities (Foster, K and others, 1995; Lynn, P, 1992; Smith, J D, 1998). None of the sport related surveys cover parents who are not members of clubs, the largest source of voluntary coaches. Some of these surveys only cover existing coaches, not athletes or other club members, the latter two being sources of voluntary coaches.

All these research programmes acknowledge that the systems are multivariate and that variables may interact and be additive. The univariate and cross tabulation comparisons used may not reveal the complexities of the system.

The reliability of some conclusions can be questioned as the survey sampling and the metrics used violate some of the theoretical assumptions underlying the statistical tests for significance employed. The surveys do not use truly random samples, even of the convenience populations that are used. Where a 100% sample was attempted, the levels of response were low and there appeared to be correlations in the characteristics of the non-responses. Some of the metrics are not parametric, yet parametric tests, such as t-tests, are used. These statistical assumption violations may account for some of the discrepancies in conclusions between studies.
A further criticism of some research on volunteering and sport is that the treatment of the problem assumed that the system was deterministic. As the systems being studied are behavioural (individuals choose to volunteer, to play sport and to coach), it is more appropriate to treat the system as probabilistic. The effects of variables can then be viewed as being those associated with probability changes.

The general acknowledgement that the social systems involved are complex and involve a number of variables and interactions of these variables suggest that a method is required that is multivariate. Hair and others (1995, p.4), citing Hardyck and Petrinovich (1977, p. 7)), sums this up:

"These methods [multivariate analysis] make it possible to ask specific and precise questions of considerable complexity in natural settings. This makes it possible to conduct theoretically significant research and to evaluate the effects of naturally occurring parametric variations in the context in which they naturally occur. In this way, the natural correlations among the manifold influences on behaviour can be preserved and separate effects of these influences can be studied statistically without causing a typical isolation of either individuals or variables."

[words in brackets added for clarification]

The methodology of this research uses statistical modelling of a voluntary decision based on the concepts of Expectancy Theory and, more particularly, Social Exchange Theory. These Theories were examined in Section 2.6. The modelling method was chosen because it is multivariate, does not require repeat measures, is probabilistic and can handle non-parametric variables, including binary variables.

Those readers not wishing to read mathematical details of the modelling should proceed to Section 6.2. in which these details have been summarised and simplified. They can then proceed to Chapter 8. Those readers wishing to follow the

STATISTICAL METHODOLOGY

Regression is a mathematical way to model situations in which a physical or sociological phenomenon (known as a Response variable) has an association with one or more independent factors (known as Explanatory variables), the strength of the association being greater than that which might have occurred by chance. In the case of this research, the phenomenon is the decision of a volunteer to enter the Coach Education Programme, the observed outcome of this decision at a point in time being the Response Variable. The independent factors, the Explanatory variables, used in this research are derived from published research. The modelling procedure used determines which Explanatory variables give the best fit to the model and, hence, have the greatest association with the Response variable. The set of potential Explanatory variables is discussed in Chapter 9.

Linear regression is the most well known form of regression modelling. It assumes that the effects of the independent factors on the phenomenon are additive. In some circumstances, there may be reasons to believe that the effects are not additive, for example, they may be multiplicative. In such cases, it may be possible to mathematical transform the true form of association into another form in which the association is additive. The resultant regression model is a model of the transformed factors and phenomenon and has to be interpreted as such. For reasons which will emerge in this Summary, such a transformation is required for this research.

Ordinary, so called Least Squares, regression requires that both the phenomenon and all independent factors can be measured on continuous scales, that is to say that all of them can, theoretically, take an infinite number of values. In this research, the phenomenon is a decision either to enter the scheme or not. The Response variable can only take one of two values. It is a so called binary variable. It is also possible
that one or more of the Explanatory Variables may only be able to take one of a limited set of values. An obvious example is Gender, another binary variable. Another example is the number of cars available, which must be a whole number, fractions being meaningless.

Regression models including variables which can not be measured on continuous scales require transformations. Regression models with transformations form a group known as generalised linear models (GLM). These are discussed in the other Sections of this Chapter.

It has been noted that the phenomenon in this research is observed in the real world as a binary variable, the individual at any observed point in time has either entered the Coach Education Programme or has not. In reality, the individual will either have made the decision or be undecided. If undecided, the individual will have an underlying probability of making the decision to enter, but some time in the future when changes in the independent factors exceed a threshold (according to the conceptual Sports Model). Probabilities are measured on a continuous scale, from 0.0 to 1.0. However, probabilities can not be directly measured, they can only be estimated.

Whilst it would be possible to ask each individual to make an estimate, the results would be of little value as the basis of estimation would vary from individual to individual and be influenced by many factors, including the knowledge, environment and thoughts of the individual at the time of estimation.

Logistic regression is one form of GLM which considers such phenomenon and allows probability estimates to be made on a consistent rationale. The Logistic is a specific mathematical shape in the form of an ‘S’. This shape has been shown to correspond to a range of biological and sociological phenomena. Initially, the independent factors have little effect on the phenomenon, the bottom gradual slope of the ‘S’. As the independent factors increase, their effect becomes more marked, the
curve in the ‘S’. Later, as the phenomenon reaches its maximum, the effects of the independent factors again reduce, the top declining slope of the ‘S’.

Chapter 7 explains, mathematically, how logistic regression transforms observed data into probabilities and into a regression model. It also discusses the assumptions that flow from the use of the logistic transformation.

One critical test of any model is how well the model fits observations from the real world. In particular, whether or not the model fits the observations better than by pure chance. It is this ability to compare with chance that gives strength to a quantitative approach. However, so far, no one measure has been developed for logistic regression that fully measures goodness of fit. Chapter 7 presents and discusses a number of measures which are used:

1. Deviance
2. Nagelkerke $R^2$
3. Proportional Change in Error, $\varnothing_p$
4. ‘C’ Statistic

6.2.1. DEVIANCE

Deviance is a measure of the variation between the observed value of the phenomenon and the value calculated from the independent factors included in the Logistic model. On the face of it, Deviance appears to be the analogy of the regression coefficient used in ordinary linear regression. However, because Deviance relates to probabilities and not to absolute values, the analogy does not hold up. Whilst absolute values can be measured in the observed world, probabilities have to be estimates. Furthermore, in this research the observed values of the probability of the phenomenon are binary (either a coach or not a coach). The distribution of these binary value probabilities is not normal, unless a very large sample is used. The overall result is that the distribution of Deviance values is of a form which is not yet known.
As a result, Deviance is only used in this research to determine whether or not a specific independent variable is included or excluded from the model. The decision is made on the basis of the change in Deviance between the two models. Values of the Reduction in Deviance are known to be distributed as Chi Square. This allows the significance of the change to be measured and the decision to include or exclude made.

6.2.2. NAGELKERKE R²

Nagelkerke R² is one of a series of measures which have been developed to be a direct analogy to the regression coefficient in ordinary linear regression. It varies between 0.0, no association, and 1.0, perfect association, but has no sign. It gives some indication of the proportion of the overall variability which is accounted for by the independent variables in the model. When the fit is poor, it is a reasonable indicator. But it becomes less reliable as the fit improves.

6.2.3. THE PROPORTIONAL CHANGE IN ERROR, Øp

Øp is one of a number of measures developed to measure the predictive ability of a logistic regression model. It takes into account the threshold value shown in the Sports model in Diagram 5.3.1. It compares the proportion of observed cases where the predictions from the model agree and disagree with the observed value (coach or not coach). It is calculated from the 4 values in the 2 x 2 frequency table of these comparisons. It can take a value between 0.0, no prediction ability, and 1.0, perfect prediction ability. All such measures so far developed have a major weakness in that they can have high values even with a model which has a manifestly poor fit. So it is necessary to measure both the predictive ability and the fit.

The aim of this research does not include prediction. The aim is to validate a conceptual model against a mathematical model and to identify which independent

STATISTICAL METHODOLOGY
factors show significant association with the phenomenon. $\Omega_{7}$ is only calculated as additional confirmation.

6.2.4. THE 'C' STATISTIC

The 'C' Statistic attempts to overcome some of the weaknesses of the other measures. It tries to measure the extent to which the model correctly assigns a high probability to those cases where the individual is a coach and to assign a low probability to those individuals who are not coaches. It takes into account the extent to which correct assignments might have occurred by chance. It can take a value between 0.5, could have occurred by chance, to 1.0, all assignments correct.

Again, the emphasis of this measure is on prediction, although it has the advantage of taking into account the threshold. It is calculated for every model, but must be viewed in conjunction with the other measures of fit.

Chapter 7 also looks at the assumptions implicit in using a logistic transformation and discusses the effects of violations.

One assumption is that the observed values for an independent factor are not clustered. This is known as Dispersion. For example, there may be fewer coaches living in some areas than in other areas. This may be just coincidence or it may be because of some feature, or features, of areas. The feature might be associated with fewer people deciding to coach or it might be associated with a reluctance to be interviewed. Whatever the reason, the resultant data is clustered. It may be possible to include a variable to explain the clustering. But sometimes the reason is not known. The same effect may occur if an important explanatory variable is left out, possibly because it has not been recognised. The problem from Dispersion tends to increase when the proportion of cases with a particular value for an explanatory variable is very low. Over or Under Dispersion can be seen by creating two way tables of the frequencies of the values of two variables. Over Dispersion may be
present if several of the cells show a zero frequency. Under Dispersion may be present if a fewer cells show very high frequencies. Both extremes violate the assumptions of logistic regression. Problems of Dispersion are most likely to occur when sample sizes are small in relation to the number of explanatory variables. The sample size required rises much faster than the number of variables. As sociological models tend to have a larger number of important variables than biological models, it is often not economically possible to take a sufficiently large sample size to avoid Dispersion.

Another assumption is that all the predictor variables are independent of each other and independent of combinations of the other variables. In reality, especially in sociological models, some degree of dependency is always present. In a multivariate model, such dependence is known as Collinearity.

Both Dispersion and Collinearity have similar effects in that they lead to underestimates of the errors of the coefficients in the models. Fortunately they do not lead to major biases in the calculated values of the coefficients. As the objectives of this research are to validate a mathematical representation of a conceptual model and to identify significant Explanatory Variables, underestimation is not a major concern as this will mainly effect the predictive ability of the model.

Measures for Dispersion and for Collinearity are used during the construction of the models. Details can be found in the next Chapter, Sections 7.4. and 7.5.

Finally, Chapter 7 explains Dummy Variable Coding. Dummy variables are needed when a predictor variable can only take a limited number of values or the values it can take are discrete numbers. Examples are Gender and the Number of Cars Available. The number of dummy variables required is one less than the number of values which can be taken. In the case of Gender, one dummy variable is required taking the value of either 0 or 1. In the case of Cars, four dummy variables may be needed for 0, 1, 2, 3, 4 and 5+. Each of these values is represented by a set of four numbers, each of
which is coded 0 or 1. Dummy variables can be used to overcome some occurrences of dispersion arising from small sample size or rare occurrence of some values. This is the case in this research. Examples of multiple dummy variables are given in Diagram 7.6.3.

The coding of dummy variables must be understood in order to interpret a logistic regression model. This is because the coefficient for each dummy variable is calculated in relation to a Reference Dummy Variable. Thus the coefficient for ‘Female’ may be calculated relative to ‘Male’ or vice versa. The coding scheme used must be known to interpret variables with more than two dummy variables, such as Cars in the example. Fortunately, the coding scheme used does not affect the overall model nor does it affect the relative importance of any one dummy variable. One scheme of coding is used in this research and is defined in a later Chapter.

6.3. Generalised Linear Models (GLM)

Expectancy and Social Exchange Theories suggest that individuals make decisions on the balance of the sum of the perceived benefits less the sum of the perceived costs. When this sum exceeds some threshold, a decision to volunteer is taken. Until the threshold is reached, the decision to not volunteer is taken. This suggests that some form of a linear model may be applicable. The simplest forms of linear models are ordinary least squares (OLS) regression and Analysis of Variance.

Both these simple linear models are part of a family now known as the generalised linear models (GLM). The objective of any GLM concerns the explanation and the prediction of a phenomenon. There is usually a strong emphasis either on explanation or on prediction, depending upon the objective of the research. In the case of this research, the emphasis is upon prediction. Explanation implies causation. In this research, any relationship found between variables will be considered as an association, not a causation.
The choice of a specific model within the GLM family is usually dependent upon three factors:

i. ability to handle all the forms of variables involved;
ii. parsimony of the number of variables involved in the chosen model;
iii. measurement of the utility of the model in predicting.

The phenomenon being researched may be represented by a single variable or by a set of variables. This variable or variables now referred to as the response variable. Previously it was known as the dependent variable (or variables). There may be one or several factors that adequately explain the phenomenon. The variables associated with these factors are now referred to as explanatory variables. Previously they were known as the independent variable (or variable). The Generalised Linear Model can deal with situations that have a single or multiple Response Variable(s) and a single or multiple Explanatory Variable(s).

In itself as a family, the GLM makes no assumptions about the distribution form of either the response or the explanatory variable(s). Both response and explanatory variable can be normally or non-normally distributed. They can be continuous or discrete variables or a mixture.

Continuous variables are those in which the variable can take on any value along a continuous scale, i.e. a scale which does not have a natural intervals or steps. Age is a Continuous scale. Sub-divisions into years, months, etc. are artificial to allow meaningful measurement. Discrete variables are those which can only take one of finite number of classes. A discrete variable can be either dichotomous (or binary), that is existing only as one of two mutually exclusive classes, or categorical, that is existing as one of more than two mutual exclusive classes. Gender is a discrete dichotomous variable. Political Affiliation is a discrete categorical variable.

In the case of categorical variables, the classes may be Ordered or Unordered, that is the order of the classes has a significance. The boundaries of the classes in an ordered
categorical variable are not set at equal intervals. For example: “Hot”, “Tepid” and “Cold” represent an ordered categorical for temperature. “Above Freezing” and “Below Freezing” represents a binary variable for temperature. The centigrade scale represents a continuous variable for temperature. “Labour”, “Liberal”, “other Parties” represents an unordered categorical variable for political affiliation.

The choice of a specific model from within the GLM family is dependent upon two factors:

i. ability to handle all the forms of both response and explanatory variables involved;

ii. the assumed form of the relationship between the response and explanatory variables.

Once a suitable GLM has been chosen, it is then necessary to choose a specific model for the sample data collected from the chosen research population. In all forms of GLM, several models are possible.

In the simple forms of GLM (for example analysis of variance and OLS regression) one unique model solution exists for every combination of variables within the total set of explanatory variables. The model fitting methods used (variance partitioning for analysis of variance, minimisation of the sum of errors for OLS regression) ensure that the single, unique solution is found.

In the case of more complex models, there may be more than one solution for any given combination of explanatory variables used. This is because the model form can not guarantee that a unique solution exists. The reason for this is shown later in this Section. A consequence is that the model fitting procedures use convergence algorithms that iterate until a ‘fit’ criterion no longer improves. Then, for a given set of explanatory variables, the solution to the fitted model may depend upon 4 factors: the initial fitted model where the convergence algorithm starts;
i. the method by which the explanatory variables are entered into the algorithm;
ii. the order in which the explanatory variables are entered into the algorithm;
iii. the ‘fit’ criterion used;
iv. the value set for the ‘fit’ criterion to stop iterations.

It is now generally accepted that the ‘best’ model for a given set of data is the one that offers:

- parsimony of the number of variables involved in explaining;
- suitable measurement of the utility of the model in predicting.

These 2 considerations determine the criterion for selection of the ‘best’ model fitting the research data. These 4 factors and 2 considerations are discussed in relation to the model type chosen for this research, logistic regression, in Chapter 7.

6.4. GLM Components

Every GLM has three components:

A random component. This is the response variable, the values of which are a random set of observations from a defined population. Most GLM’s assume that these observations are independent of each other. The response variable is assumed to be represented by its mean value with a variation around this mean approximating to a specific form of probability distribution. GLM’s exist for Normal, Poisson, Binomial and Multinomial distributions.

A systematic component. This is the set of explanatory variables plus a random variable with a fixed structure to the set. The majority of GLM assume this fixed structure to be linear. In other words, the effects of the explanatory variables are assumed to be additive and proportional. This relationship gives rise to the general equation for the multi-variable form of linear regression:

STATISTICAL METHODOLOGY
\[ \eta_i = \alpha + \beta_1 x_1 + \beta_2 x_2, \ldots, \beta_j x_j \]  
\text{(Equation 6.4.1.)}

where:

- \( \eta_i \) is the linear predictor variable
- \( x_1, x_2, x_j \) are the set of explanatory variables
- \( \beta_1, \beta_2, \beta_j \) are the coefficients (or weightings) of the explanatory variables
- is the value of \( \eta_i \) when all the explanatory variables are zero, i.e. the model constant

A link function. This controls the way in which the systematic component is mapped on to the random component. Where the random component is univariate and normally distributed, the link can be direct. This is the case with OLS regression and analysis of variance. But the link can be a function. Then Equation 6.4.1. can be expanded to a general form:

\[ g(\mu_i) = \eta_i = \alpha + \beta_1 x_1 + \beta_2 x_2, \ldots, \beta_j x_j \]  
\text{(Equation 6.4.2.)}

where:

- \( g \) is the link function
- \( \mu_i \) is the random component

The linear predictor variable, \( \eta_i \), is then converted to the random component of the model and its values become expected, rather than observed, values. This requires a computational algorithm based upon the nature of the link function and estimating the most probable values. Because the values are now probable, the estimated fitting is no longer absolute, it is a maximum likelihood estimate. Consequently, the fit is no longer absolute and unique. (Hutcheson, G, and Sofroniou, N, 1999)
There are a number of forms that a link function can take. The choice of a particular link function is partially to reflect the underlying process being modelled and partially to be able to statistically handle the distributions of the response and explanatory variables involved. The next Sections consider the nature of the response and the explanatory variables used in this research and why the choice of logistic regression was made.

6.5. The Response & Explanatory Variables

In Section 1.3., it was shown that the response variable was binary. An individual had either ‘Entered’ the Coach Education Programme or ‘Had Not Entered’ the Coach Education Programme. The effect of this study basis is that the observed values of the response variable can ONLY take one of two values: 0 or 1. Similarly, the values of the response variable predicted by the model must lie within the range of 0 to 1.

The link function for OLS regression is direct. This means that there are no upper or lower constraints on the predicted values of the response variable, that is the response variable can take any value between + and - infinity. This makes this link function unsuitable for this research.

The direct link function of OLS regression also assumes that the explanatory variables are all univariate and normally distributed. The reviews of volunteering and of sports coaching and the research surrounding these topics reviewed in Chapters 2 and 3 give many examples of potential explanatory variables that are not univariate or normally distributed. Three examples will illustrate this. ‘Gender’ is a discrete, binary, variable with a non-normal distribution. ‘Club’ is also a discrete variable that may have several categories. It will not be normally distributed and may not be univariate. Measurements of ‘Attitude’ are often on Likert scales. Likert scales are bounded in their upper and lower values. The distributions of responses to Likert scaled questions are frequently highly non-normal, sometimes not uni-variate. Again, the
natures of the explanatory variables make OLS regression (and analysis of variance) unsuitable.

The direct link function of OLS assumes that the response variable reacts in a linear fashion to the explanatory variables. A great deal of quantitative research on biological and sociological systems indicates that they do not behave in this fashion. Expectancy and Social Exchange Theories, discussed in Section 2.6, suggest that the perceived value of the net rewards of a situation are gradually cumulative towards a threshold value and then accelerate away when that threshold has been reached.

This is reflected in this research. It could be expected that most people will see little or no value in becoming a coach. Some of them will gradually see a value and hover in making a decision to become a coach. A relatively small perceived increase in the rewards may tip the balance. They will then change their minds and decide become a coach. There is an analogy with the life/death ‘S’ curve frequently encountered in biological systems. Such a curve can also be limited within the bounds of 0 and 1.

In summary, this research requires a system which:

- allows the random component to be binary;
- allows the systematic component to have a mixture of continuous and discrete variables;
- maps the systematic component on to the random component by means of an ‘S’ shaped curve with limits of 0 and 1.

6.6. Potential Forms of GLM

Three forms of GLM could be considered:

- Discriminant Analysis
- Loglinear Analysis
- Logistic Regression
All three allow the predictor variable to be placed in one of two or more classes, that is it can be a discrete variable.

Whilst discriminant analysis has been used extensively to place observations into groups determined by the explanatory variable observations, it can yield probabilities that are less than 0 and greater than 1. Discriminant analysis assumes multivariate normality. This assumption rules out the use of binary variables (Rice, J C, 1994).

Loglinear analysis is also mainly concerned with grouping predictions, in this case predicting counts in contingency tables of combinations of the response and explanatory variables. Loglinear analysis uses a log function as the GLM link component. Whilst this method can handle binary data, it can only be used when all the explanatory variables are categorical (Rice, J C, 1994). This requirement can be accommodated by transforming all the variables used in the research into categorical variables. There is some loss of information when a variable is ordered. Two examples of ordered variables in this research are literacy level and family size. There is also a practical difficulty in that loglinear analysis calculations have problems if there are sparse cells (Hutcheson, G, and Sofroniou, N, 1999). A sparse cell is a cell in a contingency table where the count is zero or very small. The required sample size to avoid scarcity rises steeply with the number of variables used and the number of categories in each variable.

Logistic regression satisfies all three considerations given at the last paragraph of Section 6.5. It is considered to give satisfactory and robust solutions to a wide variety of biological, behavioural and social phenomena (Press, J, and Wilson, S, 1978). It was developed largely for use in medicine, pharmacology, epidemiology, psychology and sociology (Hosmer, D W, and Lemeshow, S, 1989). It has been used to study volunteering in the US (Sundeen, R A, 1992) and in education (Fischer, G H, 1983. Hayden, M, and Carpenter, P, 1990). In all these studies, the logit link function was shown to be a good mathematical descriptor of the underlying
biological and social processes. The same shaped function was also found in a study of the choice of a Leisure Centre (Mintern, J H., 1994).

The features and assumptions of logistic regression are described in detail in Chapter 7 that follows. The quantitative and qualitative variables selected for use in the research are described in Chapter 9. The construction of the model using the computer package SPSS, its statistical testing and its interpretation is outlined in Chapter 10.
7. LOGISTIC REGRESSION

7.1. The Logit Function

The probability that an individual becomes a coach can be represented by the following:

\[ p = \frac{e^z}{1 + e^z} \]  
(Equation 7.1.1.)

where:
- \( p \) is the probability of case being a coach
- \( e \) is the natural logarithm
- \( z \) is the systematic component of the GLM, the left hand side of Equation 6.4.1.,

\[ \alpha + \beta_1 x_1 + \beta_2 x_2, \ldots, \beta_j x_j \]

or

\[ p = \frac{e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_j x_j)}}{1 + e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_j x_j)}} \]  
(Equation 7.1.2.)

Using some data adapted from Hutcheson & Sofroniou. 1999. (Table 4.3., Page 121) in Equation 7.1.1. with a single predictor variable \( x_1 \) and calculated \( \alpha = -2.741 \) and \( \beta = 0.4 \) results in an S shaped curve as shown in Diagram 7.1.1., known as the Logistic curve.
A probability expresses the degree of certainty of an event occurring, ranging between 0 (absolutely not) to 1 (absolutely certain).

An alternative approach is to express uncertainty in terms of the odds of an event occurring:

$$O = \frac{p}{1 - p}$$

(Equation 7.1.3.)

where

- $O$ is the odds of being a coach
- $p$ is the probability of case being a coach, as per Equation 7.1.1.

This is a specific use of the word ‘odds’ and one which is frequently used in literature on logistic regression. It will be seen from Equation 7.1.3. that ‘odds’ are measuring the ratio of the probability of being a coach to the probability of not being a coach. ‘Odds’, in this sense, measure the number of times it is more likely that the outcome is to be a coach over the outcome of not being a coach. Thus the ‘odds’ of 3.0 indicate that it is 3 times more likely for the individual to be a coach than not to be a coach.

If Equation 7.1.1. is substituted for $p$ in Equation 7.1.3., the expression is obtained:
Using the same data as for Diagram 7.1.2., the odds curve using Equation 7.1.4. is shown in Diagram 7.1.2. This shows how the odds of being a coach increase with the systematic component. The odds represent the number of times that the event is more likely to occur over the base situation, not being a coach.

The functions in Diagram 7.1.1. and Diagram 7.1.3. are both non-linear. But if the natural log is taken of the odds, a linear function is obtained. The natural log of the odds is known as the Logit.

Natural Log (odds of being a coach) = Logit(c)  

Logit(c) = a + β₁X₁ + β₂X₂ + ... + βₙXₙ  

EQUATION 7.1.5.  

EQUATION 7.1.6.
Again, using the same data as for Diagram 7.1.2., the logit function produces a straight line as shown in Diagram 7.1.3.

**Diagram 7.1.3.**

**THE LOG ODDS or LOGIT FUNCTION**

(Hutcheson, G, and Sofroniou, N, 1999).

The use of log odds thus allows a form of regression analysis to be used. This is logistic regression. There is a linear additive increase with an increase of the systematic component. As the systematic component increases by 1 unit, the log odds increase according to the parameters of the fitted model. In the illustration, as the systematic component increases by 1 unit, the log odds of being a coach rises by 0.4 units ($z$) with a constant component of -2.741 ($\alpha$).

It will be seen that the major difference to OLS regression is one of interpretation. The units in the logistic regression model are log odds, not the absolute units of the response and predictor variables. This has particular implications in the interpretations of the estimates of the coefficients and their errors. This is discussed in Section 10.8.
7.2. **Assumptions**

In contrast to OLS, logistic regressions makes few assumptions. Three that do apply are:

1. The response variable is a random binary variable following the binomial distribution.

2. The explanatory variables:
   - are independent of each other
   - are measured with little or no error
   - have constant variance
   - are all present

3. The additive effects of the explanatory variables are adequately mapped by the logit function

Each of these three assumptions is now discussed in the context of this research.

7.2.1. **RANDOM RESPONSE VARIABLE**

The assumption of randomness in the response variable is largely dependent upon adequate definition of the population and the adequacy of the sampling of that population. Both these topics are discussed in Chapter 8. There is also the implied assumption that the two 'states' used for the binary response variable are defined in such a way that they are mutually exclusive. This condition is satisfied in this research as an individual has either entered or not entered the Coach Education programme and this can be verified from the list of qualified coaches.
7.2.2. EXPLANATORY VARIABLE ASSUMPTIONS

The assumptions about the explanatory variables are more difficult to fully meet. Indeed it is unlikely that all of them will be met in a sociological setting.

7.2.2.1. Independence

Independence is the most important of these assumptions. Independence can be measured in terms of collinearity and this is described in Section 7.5.

7.2.2.2. Errors

It almost certain that the measurements of most explanatory variables will be with error, especially where the variable concerns an attitude or is a surrogate for some unmeasurable, possibly complex, underlying variable or variables. The effects of including variables with significant errors are largely on the accuracy of the estimates of the coefficients. Providing that the variance is constant, the estimate of the coefficient will be unbiased. So the incorporation of variables with significant errors mainly effects the 'statistical' significance of the estimates of the coefficients, that is whether or not they arose by chance. This research is more concerned with 'substantive' significance, that is the practical importance of a variable and it's effect relative to other variables (Hutcheson, G, and Sofroniou, N, 1999).

7.2.2.3. Variance

When a variable has constant variance, this is known as homoscedascity. It can be seen when the variable is plotted against another, related or second, variable. The variations lie in a parallel band of association between the two variables. This is illustrated in Diagram 7.2.1.
When the variance is not constant, known as heteroscedasticity, the variability does not lie in a parallel band. The most frequent type of heteroscedasticity occurs when the variance increases as the values increase. This is illustrated in Diagram 7.2.2.

Other forms of heteroscedasticity can exist, for example causing the band to be a balloon or a dumbbell.
Where there is more than one explanatory variable or where the response variable is binary, the screening process illustrated in Diagrams 7.2.1. and 7.2.2. can not be used. This is the case with logistic models. In these cases, the test for homoscedascity has to be made by plotting the model residuals against the fitted values of the response variable.

Heteroscedascity can occur for 2 reasons.

- The variance of one or more explanatory variables is not constant. This frequently occurs when the explanatory variable has a multimodal distribution.
- The relationship between the response variable and one or more explanatory variables is not linear.

One method of rectifying heteroscedascity is by transforming the non-constant variable. The transformation may be a simple mathematical one, for example log or inverse, when the variable is unimodal. Alternatively, a continuous non-constant variable can be transformed into a discrete variable so that the variance within each category is, relatively, constant. This is particularly useful when the original variable is multimodal. There are two disadvantages to using these transformations:
• Increased difficulty in practical interpretation of the model coefficients.
• In the case of a transformation of a continuous variable to a discrete variable, a loss of information.

Presence of heteroscedasticity is not fatal. It will decrease the power of prediction of the model as the coefficients may be biased and testing of their significance may be effected by poor estimation of their standard errors. Once again, the weaknesses from heteroscedasticity will not be important to this research as the emphasis is on substantive and not statistical significance.

7.2.2.4. Completeness

As there are no well researched theories on the decisions of adult volunteers to enter vocational training in sport and as some sociological variables, especially attitudes, are difficult to measure, it is likely that one or more explanatory variables will be omitted. The extent to which omission effects the results can be judged by analysing the residuals.

7.2.3. THE LINK FUNCTION

It has been shown that the logit function is applicable to many social models and that the results of logistic regression are not significantly effected by the choice of the link function (Aldrich, J H, and Nelson, F B, 1984).

7.3. Goodness of Fit

As the Logit function (Equation 7.1.4) is based upon probabilities and as these probabilities must be estimates, the true values being unknown, the normal least squares method used with ordinary linear regression can not be used to solve the simultaneous equations for the coefficients. Furthermore, the variance of the
observed response variable (Decision) is not constant. Instead it depends upon the true and unknown probability. In addition, the distribution of the binary response variable is only normal if the sample size is very large. (Collet, D, 1991).

The most common statistic for measuring ‘goodness of fit’ is the Deviance. This is the main statistic provided in SPSS, the computer package used in this research. It is defined as:

$$D = -2 \log \left( \frac{L_c}{L_f} \right) = -2(\log L_c - \log L_f)$$  
(Equation 7.3.1)


Deviance is often referred to as -2LL, a short hand for the Equation 7.3.1., -2 * Log Likelihood.

where:

$$D = \text{Deviance. The extent to which the current model deviates from the full model.}$$

$$\hat{L}_c = \text{Maximum Likelihood under the Current Model. The Current Model includes the unknown coefficients set to their Maximum Likelihood Estimates.}$$

$$\hat{L}_f = \text{Maximum Likelihood under the Full or Saturated Model. The Full Model includes the same number of parameters as there are observations and fits the model perfectly.}$$

$$D \text{ for binary data can be shown to be:}$$

$$D = -2 \log \sum_j \left\{ \hat{p}_j \log \hat{p}_j + \log(1 - \hat{p}_j) \right\}$$  
(Equation 7.3.2.)

(Collett, D, 1991).

LOGISTIC REGRESSION
Where:

\( \hat{p}_j \) = is the fitted probability for observation j.

As the deviance for binary data is ONLY dependent upon the fitted probabilities, it can not measure anything about the fit of the observed data to the calculated probabilities. As a consequence, it is not an informative metric for 'goodness of fit'. (Collett, D, 1991).

There is another problem with the use of deviance. To test the significance of deviance, a distribution of deviance must be assumed. The distribution normally used in literature has been chi-squared. Recently, a number of authors have shown that deviance for binary data is not distributed as chi-square. Furthermore, a requirement of a chi-square is that more than 75% cells used in the test must contain 5 or more observations. In the case of the binary data being considered, the number of cells is related to the number of covariate patterns produced by the data set. In turn, the number of covariate patterns is related to the number of variables and the nature of the variables. Binary variables produce fewest patterns, categorical variables more patterns and continuous data even more patterns. Thus a data set with only a few continuous variables can produce a very large number of patterns. Even with large data sets, this may result in many cells with zero observations and more than 25% with less than 5 observations. In such cases, and this research falls into this category, significance testing of deviance using chi-square is not valid. To-date research has failed to find a solution to this problem (Collett, D, 1991; Hosmer, D W, Jnr. and Lemeshow, S, 1999). The measurement of fit of a model will be returned to shortly.

Change in Deviance, however, is used to determine if an iteration has improved the model and will be used in this research to determine if an explanatory variable is included or is excluded from the model. Change in Deviance measures the improvement of an initial model and requires comparison of two models.
At the extreme, one model would be the model fitting the data only with an intercept and not including any of the independent variables. This is similar to a linear regression with any effects of the independent variables being assumed to be part of the error term. Another model is one fitted with an intercept and one or more of the independent variables. The fits of these two models can be compared, the second model having an improved ‘fit’ due to the presence of at least one additional independent variable.

In each case the Deviance, $D_0$, is calculated (Equation 7.3.1.).

$D_0$ is the deviance of the model with the intercept only, referred to in SPSS as “Initial Log Likelihood function - 2 Log Likelihood.” Equivalent to the total sum of squares in linear regression.

$D_m$ is the deviance of the model with the intercept and ‘m’ variables, referred to in SPSS as “2 Log Likelihood, Chi Square.” Equivalent to the unexplained sum of squares in linear regression.

Then:

$$G_m = D_0 - D_m$$

(Equation 7.3.3.)

where:

$G_m$ is the Reduction in Deviance between these two models, referred to in SPSS as “Model Chi Square.” Equivalent to the regression sum of squares.

(Menard, S, 1995.).

Unlike Deviance alone, Changes in Deviance are approximately distributed as chi-square. This allows statistically significance to be measured. The degrees of freedom are determined by the number of independent variables in the larger model minus the number in the smaller model.

LOGISTIC REGRESSION
are equal to the difference in the number of variables between the two models. There is a proviso. One model must either include one or more additional variables to the other model or exclude one or more of the variables of the other model, i.e. one model must be nested in the other model. This means that the significance of the interaction effect between two variables can only be tested against a model that includes at least both the variables. But the distribution of the deviance differences does allow a method of testing whether or not one or more additional variables improve the model fit.

Whilst Change in Deviance allows decisions to be made on the addition or exclusion of variables, there have been no entirely satisfactory overall measures of fit. A number of alternatives have been suggested for Deviance. But, to-date, none of them fully satisfy the conditions or assumptions of logistic regression. In this research 3 statistics have been used as ‘indicative’ of fit, each interpreting ‘fit’ in a different way:

- The Nagelkerke R², which attempts to measure the proportion of the variation explained by the explanatory variables included in the model. This is an equivalent to the Pearson Correlation Coefficient, R² of OLS linear regression (SPSS Inc. 1999).
- The Proportional Change in Error measure, ϕp, a measure of the predictive ability of the model (Menard, S. 1995).
- The ‘C’ statistic that attempts to measure the extent to which the model assigns high probabilities to cases that actually have the positive outcome, being a coach and low probabilities to those cases that actually have the alternative outcome, not being a coach. It ranges over a scale of 0.5, if the assignments could be by pure chance, to 1.0, if the model always assigns a higher probability to positive outcome cases than to alternative cases (SPSS Inc. 1999, citing Hanley, J A, and McNeil, B J, 1982)

The ‘C’ Statistic is derived as follows. A cross-tabulation table is constructed of the values of the Predicted variable (probability of being a coach) against their actual state (i.e. 1 or 0 according to whether they were or were not coaches). To enable such a

LOGISTIC REGRESSION
Each pair of resultant ordinal values is then compared. Those pairs with the same actual state are eliminated (i.e. pairs both coaches or pairs both not coaches).

The probability of the remaining pairs having different states by chance alone is then calculated using Somer’s d Statistic. Somer’s d lies between -1 and +1. 0 indicates that there is no association between the pairs, i.e. where they have the same ordinal value, this could have occurred by chance.

This research is concerned with a person deciding to become a coach, not with those not becoming a coach, that is those that had a positive outcome. We wish to know if a positive outcome is associated with the predictor variables and if such an association occurs by more than chance. The strength of such an association may be measured by the ‘C’ Statistic. The ‘C’ statistic is derived from Somer’s d as follows:

\[ C = \frac{d}{2} + 0.5 \]  
(Equation 7.3.4.)

Where

\[ C = \text{‘C’ Statistic} \]
\[ d = \text{Somer’s d} \]

The ‘C’ Statistic is sometimes referred to as a ‘guide to goodness of fit’. It provides percentages of the cases that are purported to be ‘correctly’ and ‘incorrectly’ classified by the fitted model. Whilst the classification table indicates the ‘success’ of the model in prediction, the classification table does not indicate how adequately the model fits the data. This is because each observation is assigned to either 1
('success') or 0 ('failure') according to whether or not the predicted probability of that observation is above or below some threshold probability. This threshold must be predetermined by the researcher. The default value in computer programmes is set at 0.5.

The following illustrates the problems of using and interpreting classification tables.

Table 7.3.1.

<table>
<thead>
<tr>
<th>Observed:</th>
<th>Predicted</th>
<th>% Correct Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>Not Coach</td>
<td>137</td>
<td>14</td>
</tr>
<tr>
<td>Coach</td>
<td>37</td>
<td>44</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.3.1. could be interpreted as follows. The model gives a 90.77% correct prediction of individuals who are not coaches and a 54.32% correct prediction of individuals who are coaches, overall, giving an 77.92% correct set of predictions.

However, this assumes that a correct value has been placed upon the threshold. Computer packages set a default of 0.5, although they offer the ability to alter this default value. Using the value of 0.5 implies an assumption that a predicted probability of 0.5 or more places that case in predicted 'success' category. In the case of this research that individual is most likely to be a coach, i.e. the predicted value of the binary dependent variable, y, is 1. Similarly, if the predicted probability is less than 0.5, the assumption is that the predicted binary value of the dependent variable is 0.

The data used to construct Table 7.3.1 is illustrated in Diagram 7.3.1. with the threshold set at 0.5.
Diagram 7.3.1.

DATA FOR TABLE 7.3.1.

In many situations, including this research, there is no theoretical reason for assigning any specific value to the threshold. If the threshold in the above example is increased to 0.7, the classification table then becomes as per Table 7.3.2.

Table 7.3.2.

REVISED CLASSIFICATION TABLE (THRESHOLD 0.7)

<table>
<thead>
<tr>
<th>Observed:</th>
<th>Predicted</th>
<th>% Correct Prediction =</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>Not Coach</td>
<td>147</td>
<td>3</td>
</tr>
<tr>
<td>Coach</td>
<td>37</td>
<td>45</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The model then gives a 98.00% correct prediction of individuals who are not coaches and a 55.57% correct prediction of individuals who are coaches, overall, giving an 83.12% correct set of predictions. The overall and non-coach prediction percentages improve, apparently, quite a lot, whilst the coach predictions rise, apparently, only marginally. These effects are only due to changing the threshold value and to frequency distributions of two groups. The calculated probabilities remain the same. Diagram 7.3.2., illustrating this, is only a repeat of Diagram 7.3.1. with the threshold moved.
Using the data plots alone, arguments could be made for the threshold to have other values, with consequent changes in the contingency table. An estimation of any real threshold value can only be made using additional evidence.

The use of a single threshold value across the entire data set also assumes that there is a single threshold for all cases. In the current research, the assumption is that every individual has the same threshold for making the decision to become a coach and, furthermore, that this threshold is independent of all other factors, including the independent variables. Put another way, a single, fixed threshold implies that not only will all individuals make the decision at the same threshold, it implies that the threshold does not even change with time. The latter is unlikely as this would, in turn, imply that the personal perceived values of gains and losses in making the decision are fixed and not effected by changing social values and expectations of peer groups.

The practical value of classification tables and the predetermined thresholds can only be judged in conjunction with a plot of all the predicted probabilities. If the majority of observations are grouped near to either 1 (if observation is ‘coach’) or 0 (of the
observation is ‘not coach’), then the table does indicate, but not measure, the
‘goodness of fit’ and the value of the threshold is not crucial. If a significant number
of observations lie on the ‘wrong’ side of the threshold, this may indicate that the
threshold is incorrect. If many observations are grouped around one value between 0
and 1, this may indicate that the model is probably not very good at assigning the
observations. However, it may also mean either that the independent variables
included account for only a small proportion of the variability or that the sample size
is too small to calculate accurate values for the coefficients. The tables becomes less
valuable when a significant number of observations lie scattered between 0 and 1.

7.4. Dispersion

At the heart of logistic regression is the calculation of odds:

\[ O = \frac{p}{1 - p} \]  
(Equation 7.4.1.)

and the use of the logit function, being the log odds (Equation 7.1.3.)

The estimation of probability is made from the observed counts of the response
variable in relation to one or more explanatory variables. This process can be
visualised as contingency tables. Below is a contrived example to illustrate.

<table>
<thead>
<tr>
<th>Birth Area</th>
<th>Coach</th>
<th>Not Coach</th>
<th>Total</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inner City</td>
<td>7</td>
<td>15</td>
<td>22</td>
<td>0.32</td>
</tr>
<tr>
<td>City Suburb</td>
<td>17</td>
<td>55</td>
<td>72</td>
<td>0.24</td>
</tr>
<tr>
<td>Urban</td>
<td>5</td>
<td>40</td>
<td>45</td>
<td>0.11</td>
</tr>
<tr>
<td>Rural</td>
<td>2</td>
<td>20</td>
<td>22</td>
<td>0.09</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>130</td>
<td>171</td>
<td>0.18</td>
</tr>
</tbody>
</table>

**Table 7.4.1.**

**BIRTH AREA & PROBABILITY OF BEING A COACH**
Then the odds for each cell of Table 7.4.1. can be calculated using Equation 7.1.2.
where \( p = \frac{\text{the value "Coach" divided by the value "Total"}}{} \). All the odds entries lie between 0.0 and 1.0 as all the values of \( p \) are positive and greater than 0.

But supposing there were no observations of ‘coach’ from ‘Rural’ birth areas. That cell “Coach” in Table 7.4.1. would be ‘0’. The calculated odds for this cell would then be 0.0 and the logit would be infinitely small. Similarly, if all observations for ‘coach’ were in the ‘Inner city’ cell, the calculated odds of that cell would be 1.0 and the logit would be infinitely large. In both cases, there may be calculation problems in the logistic regression model fitting. Calculation problems may also occur if there are many cells with relative low counts. These will to very small calculated odds and, hence, logits approaching 0. These effects are known as over and under dispersion.

Extreme dispersion (over or under) may occur for several reasons, of which the following are the most common:

i. The set of observations may be from a non-random sample

ii. The sample may not be large enough, relative to the number of explanatory variables and their categories

iii. The distribution of the response variable does not correspond to a binomial distribution.

iv. A significant explanatory variable has been omitted

v. The scale or number of categories of one or more explanatory variables may be too detailed (or few) or too large (or many). In the former case, the data may be concentrated in only a few cells. In the latter case, the data may be too thinly spread do that many cells have no entries.

Reasons (i) to (iii) can be avoided by ensuring that the sampling is random, that the sample size is adequate and that the response variable is defined in such a way that the two categories are mutually exclusive and cover all cases. Sampling is discussed in Chapter 8. The definition of ‘Coach’ was discussed in Chapter 4. The research
attempted to minimise the risk of omitting a significant explanatory variable (Reason (iv)) by reviewing the literature surrounding volunteering and coaching in Chapters 2 and 3.

The full extent of problems associated with reason (v), explanatory variable scales, will not be known until after data collection has finished. Potential dispersion problems can be identified by creating contingency tables of the observed counts of the data from every pair of discrete variables, including the response variable. Potential problems may arise if either there is a cell with zero count or more than 20% of cells have counts of less than 5. This is the same test that is applied to the Chi Square test (Black, T R, 1999).

The method chosen in this research to rectify such potential problems are to collapse categories in the variable causing the problem so that not less than 20% of cells have counts less than 5.

A similar situation can arise if the observed distribution of a continuous explanatory variable is multinomial, that is the variable has several values with high frequencies and other values with low frequencies. This can occur with attitudinal variables where a particular attitude is held by a specific sub set of the population. It may also occur through deficiencies of the random sampling. In this research, the continuous variable will be transformed into a discrete variable and the rule in the previous paragraph applied. Categories for the discrete variable will be defined so that the peak frequencies will be at the centres of chosen categories.

In the extreme, if the number or severity of dispersion problems becomes too great, it may be impossible to calculate a solution to the model. If calculation is possible, but there is large over or under dispersion, the estimated coefficients and their errors over estimated. The model will not be reliable for prediction or statistical testing of hypotheses. Providing the under or over dispersion is not too great and a model can

LOGISTIC REGRESSION
be calculated, the model can be used for purposes of substantive significance, as is the case in this research.

The reason for the errors caused by over or under dispersion lies with the assumption that the response variable is binary and has a binomial distribution of errors. The binomial distribution has a variance that is a function of the mean of the distribution. Consequently, if the variance from the explanatory values is greater than the assumed value (over dispersion) or less than the assumed value (under dispersion), the assumption regarding the variance of the response variable will be violated and subsequent calculations will be incorrect.

The degree of dispersion of a fitted model can be estimated as follows:

\[ s = \frac{D_m}{df} \]

(Equation 7.4.2.)

where:

- \( s \) is a constant
- \( D_m \) is the Deviance of the model (See Section 7.3.)
- \( df \) is the number of degrees of freedom of the model

If the data is over dispersed, \( s > 1 \). If the data is under dispersed, \( s < 1 \).

The constant, \( s \), can be used to correct the estimates of the standard errors of the model coefficients. This is important if hypotheses are to be tested involving use of the estimated standard errors. These include testing that a coefficient significantly differs from 0 or that the coefficient of one variable significantly differs from that of another variable. The correction using the constant, \( s \), from Equation 7.4.2. is:

\[ \text{ASE}_j = \text{SE}_j \times s \]

(Equation 7.4.3.)
where:

\[ \text{ASE}_j \] is the Adjusted Standard Error for explanatory variable j

\[ \text{SE}_j \] is the estimated Standard Error for explanatory variable j

(Hutcheson, G, and Sofroniou, N, 1999).

7.5. Collinearity.

Also known as Multicollinearity. Collinearity exists when two or more explanatory variables are significantly correlated. When the correlation is perfect, a singularity is said to occur, that is to say that all the values of one explanatory variable can be obtained from the values of another explanatory variable. It is also possible for one explanatory variable to be perfectly explained by two or more other explanatory variables.

Most computer programs for logistic regression, including SPSS, check for singularities and high correlations between pairs of explanatory variables. If a singularity is found or a pair correlation with a squared coefficient greater than a set amount (usually 0.80), the programme will not proceed and will issue a warning. No known programs check directly for multicollinearity. However, SPSS and other programs allow a separate calculation of the statistic Tolerance:

\[ T_j = 1 - R^2_{mj} \]  

(Equation 7.5.1.)

where:

\[ T_j \] is the Tolerance statistic for explanatory variable j

LOGISTIC REGRESSION
$R^2_{mj}$ is the squared Multiple Correlation Coefficient between explanatory
variable j and all the other explanatory variables in the model.

Then:

$$R^2_{mj} = 1 - T_j$$  \hspace{1cm} \text{(Equation 7.5.2.)}

Tolerance statistics were computed in this research within the SPSS package using
the Regression and Linear commands available within the Analyze command.

The effect of including explanatory variables with collinearity is that the estimated
coefficients of the model are too high, that is the effect of an explanatory variable is
exaggerated. Collinearity also leads to high estimates of standard errors. This will
restrict the ability to test for statistical significance.

It is generally accepted that Multiple Correlation Coefficients less than 0.80 do not
introduce undue bias and that Multiple Correlation Coefficients greater than 0.90
must be corrected. Alternatively, using Equation 7.5.1., Tolerances greater than 0.20
are acceptable and action must be taken where Tolerances of less than 0.10 require
are encountered (Menard, S, 1995; Hutcheson, G, and Sofroniou, N, 1999).

If significant collinearity is encountered, there are three alternatives available:
i. The sample size is increased by additional sampling. A larger sample reduces the
effects of collinearity as it reduces the standard errors of the variables. In this
research, additional sampling will not be feasible due to time and funding
constraints. Also retrospective sampling may introduce additional variables as a
result of the time elapsed between the original sample and the second sample.
During the elapsed time, significant events may have occurred. In addition, during
the elapsed time, individuals interviewed in the original sample may talk to and
influence the responses of individuals interviewed in the second sample.
Explanatory variables highly correlated with each other can be combined into a new, single explanatory variable. To be of any practical use, the combined variable must measure a concept that can be related to reality and there must be some theoretical basis for the means of combining the variables, e.g. other published research or empirical reasoning indicates that the effects of the two original variables are additive.

iii. An explanatory variable causing high collinearity (i.e. with a high multiple correlation in respect of the other explanatory variables in the model and the response variable) can be removed from the model. However, such removal may cause additional problems, especially if the removed variable has, in its own right, a large effect on the response variable. Removal can mean that the model lacks a significant variable thus violating one of the assumptions of logistic regression (see Section 7.2.2.4.).

(Hutcheson, G, and Sofroniou, N, 1999)

A further effect of collinearity can be caused by the mechanism used to enter explanatory variables into a model. This can be illustrated by the Venn diagrams used in the following illustration. The variances of the response variable, Y, and two explanatory variables, X₁ and X₂, are represented by 3 circles. Explanatory variables X₁ and X₂ exhibit collinearity with respect to response variable Y and to each other. Both circles X₁ and X₂ overlap circle Y indicating that both these explanatory variables contribute to the variance of the response variable. In addition, circle X₁ overlaps circle X₂ indicating that the two Explanatory variables are also related to each other to the extent of their shared variances.

In Diagram 7.5.1., X₁ and X₂ are introduced simultaneously to the model. The variance of Y attributed by the model to X₁ is hatched horizontally. The variance of Y attributed by the model to X₂ is hatched vertically. The areas of Y not hatched are not attributed to X₁ or X₂. This includes an area of overlap between X₁ and X₂, their
shared variance. The model does not attribute this area either to $X_1$ or to $X_2$. The unhatched overlap area in Diagram 7.5.1. represents this unallocated variance.

Diagram 7.5.1.
SIMULTANEOUS INTRODUCTION

Alternatively, $X_1$ could be introduced to the model first and then $X_2$ introduced to this new model already inclusive of $X_1$, a sequential introduction. The effects are shown in Diagram 7.5.2.
Because explanatory variable $X_1$ is now introduced before explanatory variable $X_2$, the shared variance area in Diagram 7.5.2. is attributed by the model to explanatory variable $X_1$. As a result, explanatory $X_1$ in Diagram 7.5.2. now appears to be contributing more of the variance of the response variable $Y$ than it did in Diagram 7.5.1.

Explanatory variable $X_1$ in Diagram 7.5.2. increases in importance in relation to explanatory variable $X_2$ between the model in Diagram 7.5.1. and the model in Diagram 7.5.2. Furthermore, the explained variance of the response variable $Y$ increases between the model in Diagram 7.5.1. and the model in Diagram 7.5.2. In reality, there is no difference between the model in Diagram 7.5.1. and the model in Diagram 7.5.2. The differences are purely due to the mechanism used to construct the model (Tabachnick, B G, and Fidell, L S, 1997).

The area of the response variable, $Y$, attributed to an explanatory variable, $X_i$, is reflected in the estimated coefficient of $X_i$, in the model. Models can be built by either simultaneous introduction of variables or by sequential introduction. When collinearity exists, the method of model building will effect the model estimates. Both simultaneous and sequential introduction of explanatory variables are available in SPSS.
7.6. **Dummy Variables**

This is the process by which discrete variables, including the response variable, are coded for entry into a logistic regression model.

A discrete variable can not be entered in its original form as it is not measured on an absolute mathematical scale. It can only be represented by an artificial variable or variables. The simplest example is the variable ‘Gender’ with two mutually exclusive categories ‘Female’ and ‘Male’. These can be artificially represented as ‘0’ and ‘1’ respectively. But this is an arbitrary assignment. It can just as well be ‘1’ and ‘0’ respectively as there is no significance in the order ‘Female’ and ‘Male’. The assignment could equally well be ‘1’ and ‘7’ as there is no significance in the scale distance between ‘Female’ and ‘Male’.

The lack of significance and of scale becomes evident when an ordered multi category discrete variable is considered, such as ‘Warmth’ with categories of ‘Cold’, ‘Tepid’, and ‘Hot’. These could be coded as ‘0’, ‘1’, and ‘2’ respectively. But the introduction of scale adds spurious and misleading information as it implies that ‘Tepid’ is 1 unit greater than ‘Cold’ and 1 unit less than ‘Hot’, etc.

The lack of significance of order as well as scale becomes more evident when an unordered multi category discrete variable is considered, such as ‘Voting’ with categories of ‘Conservative’, ‘Labour’, ‘Liberal Democrat’ and ‘Other’. These could be coded as ‘0’, ‘1’, ‘2’ and ‘3’ respectively. Not only does this imply scale, it also implies order, both of which are spurious added information.

Dummy coding overcomes these problems of implied scale and order. It considers every discrete variable as a matrix of all the categories in the variable. In the case of a binary variable, the matrix consists of a pair only, e.g. ‘Female’ and ‘Male’. The three examples previously used could be dummy coded as follows
It will be seen in Diagram 7.6.3. that one category is coded ‘0’ for every dummy variable. This is known as the reference category. In Diagram 7.6.3. the reference categories are the last listed: ‘Male’, ‘Hot’ and ‘Other’. This is the coding method used as the default in SPSS and is used in this research.

When the discrete variable is entered into the model, it is the matrix of dummy variables that is entered. In the above example, there is one dummy variable for ‘Gender’, two dummy variables for ‘Warmth’ and 3 dummy variables for ‘Voting’. The model estimates coefficients for each of the dummy variables entered. This means that there is one coefficient for each of the categories except the reference category, in the above example the first category listed. A coefficient then indicates the effect of its dummy variable category to the reference category. This operates in the model as a coefficient is multiplied by the appropriate coding in the matrix. Only one of the appropriate codes is 1. All the others are 0. In the case of the reference category, all its codes are 0. Thus a coefficient is unnecessary for the reference category.
8. DATA COLLECTION

8.1. National Population

The Sports Council estimated that, in 1997, there were 2,300,00 participants in athletics in the UK. But the vast majority of these participants were involved in running for recreation (including fitness) rather than for competition. These recreational athletes are not usually associated with coaching and were not considered in this research. In the same year, 1997, it was estimated that there were 1,700 athletic clubs in the UK with 177,317 affiliated members (The Sports Council, 1996(a)).

A large proportion of these members (up to 50% in some clubs) are under the age of 18 and not eligible to become coaches. This group, under 18, are not considered in this research.

Every Under 18 club member has one or two parents or guardians. There may be 100,000 of these adults who are associated with athletics through their children, but not formally members of a club. These parents and guardians may only become a member of a club as a result of making the decision to become a coach. This group, parents of Under 18 members, is included in the research. The club records allow these parents to be identified by the surname, address and telephone number listed for each child.

These two groups (Members of 18 years of age and over plus parents and guardians of members Under 18) form a potential athletic coach population of up to 200,000. Within this population, it is thought that there are about 8,000 qualified athletic coaches in the UK (Ward, T, 2000). The exact figure is not known as there is no national register.
Even if it was feasible to sample this entire population, there are problems of
definition. Membership lists are notoriously unreliable. Lists only record those
associated with the club who have met certain criteria. Some of these criteria may
vary from club to club and, from time to time, within a club, e.g. honorary
membership, family membership, etc. Lists do not fully record individuals who are
informally associated, including parents, although many of these may be active in
keeping the club running or in supporting the athletes and officials.

The criteria for inclusion or retention on a membership list usually include reference to
payment of annual membership fees. Cessation of membership is formally covered by
the rules of each club, which may differ from club to club. Also, for administrative
and social reasons, names may not be removed from membership lists immediately
after formal cessation of membership. Because administration is voluntary, there can
be delays between election to membership, payment of fees and inclusion in and
removal from the membership list.

There may also be external reasons for either restricting or enlarging the numbers on
the membership lists. Causes can include the levying of affiliation fees by league,
county and national bodies, based upon bands of membership size, and decisions by
these bodies on the status of the club in competitions and leagues.

Thus a rigorous definition of the membership of even ‘listed’ members is neither
complete nor universal nor consistent over time.

Membership lists do give some ability to contact the parents or guardians of the
Under 18 club members. In this research, potential coach parents or guardians are
assumed to be those who are not formally members but have one or more children
under the age of 18 who have been members for at least one year. 18 was selected as
this is the minimum age at which an individual could enter the coach education
programme. One year’s membership was selected as the coach education programme
mainly runs through the autumn and the winter whilst most children join in the spring.
and summer. Parents are unlikely to know of or be able to enter the programme until the year following initial membership.

Any attempt to sample over the entire national population would have to take into account the factors associated with the differences described in this Section. Each factor would have to be represented by one or more additional explanatory variables. Each additional explanatory variable increases the sample size required and may not significantly increase the power of the research to answer the research question posed in Section 1.1.

Personally interviewing a sample drawn from over the whole of the UK was neither physically nor financially viable. The data would have had to be collected over the telephone or postal questionnaire. Postal and telephone responses can have a low response rate and may not representative of the sample chosen as gender, age, ethnicity, geographic and socio-economic factors effect response rates. Both postal and telephone interviewing tend to restrict the number of explanatory variables that can be obtained as the response rate and qualities of response can decline if the interview is too lengthy.

For these reasons, use of the national population as the sampling frame was rejected for this research. Instead a decision was taken to restrict the population used for the research. It was accepted that such a sample would severely restrict the results of the model being extrapolated to the national population. However, such extrapolation was not the prime intention of the research. The prime intention of the research is to show that a substantive Logistic regression model can be constructed using a set of data that can be readily obtained from individuals associated with athletics.

Successive surveys of volunteers in different countries have highlighted that a main reason why people volunteer is because they are asked to do so (Lynn, P, and Smith, J D, 1991). When, who and how individuals are asked is likely be conditioned by the structure and active membership of a club. The structure of a specific club is likely to

DATA COLLECTION
be a result of its location and history. These will effect the characteristics of the membership, especially the active members. The membership will be influenced by the catchment area and by the attraction of the club in terms of facilities and perceived sporting opportunities. All these may effect the number of qualified coaches in the membership. The individual coaches may positively influence the membership of the club by personally recruiting or attracting athletes with certain characteristics and by increasing their retention. Coaches may also have negative influences on club membership for the same reasons. As a result, variable ‘club’ was included in the model.

For practical and economic reasons, this research was restricted to two similar athletic clubs. This also reduces the number of variables which would otherwise have to be included to take into account. Minimising the number of variables was important in that it minimised the sample required and, hence, the costs and resources. This can be seen when the similarities of the two clubs are considered:

- from towns in the same County in South East England;
- of similar size (about 100,000);
- both based in towns serving a number of small semi-rural villages; with similar athletic facilities: an all weather, floodlit 400M track with full facilities;
- catering for all ages, both genders, all abilities including recreational and veteran, and for road running (but not race walking);
- situated in similar socio-economic areas; with a similar membership size (about 300); and of a similar age (founded about 50 years).

This decision to restrict the sampling population is justified on the basis that the objective of the research is to show the validity of a logistic regression model. The objective was not to find the entire set of variables for athletics throughout the UK or measure their relative importance.

The administration and coaching of both clubs is, and always has been, entirely by unpaid (except for expenses) volunteers, all of who are club members. Neither club

DATA COLLECTION
receives any major external funding or sponsorship and is self financing. Both clubs use local authority owned tracks. In one case, the club manages the facility under a lease from the local authority. This facility is associated with a small number of outdoor football pitches managed and hired by the local authority. In the other case, the club pays for use of the facility that is commercially managed as part of a leisure centre.

This next Section, 8.2, describes how sampling was conducted within this two club population and how the resultant sample data is related to this population.

8.2. Research Population

The two clubs provided their official membership lists at the beginning of 1999. These referred to the membership during the calendar year 1998. In each case the club indicated which members were considered to be coaches. Surrey County Athletics Association (SCAA) supplied lists of active, qualified athletics coaches as of 1998, showing the club to which each coach was affiliated. The club and SCAA lists were compared. There were cases of coaches listed in the SCAA list who were not listed as coaches by their club. In some instances, they were listed as members. For the purposes of this research, it was assumed that they were qualified coaches. In other instances, coaches listed by SCAA were not listed as members of either club. For the purposes of this research, it was assumed that these coaches had either ceased activity or had moved out of the area. They were not included in the research.

These club membership lists were then sorted into those born before and those born in or after 1981, i.e. those under and over 18 at the beginning of 1999. The former group comprised the child member population that was discarded for this population, not being eligible to be coaches. The latter group comprised the adult members used in the research. These were then divided into those known to be coaches and the non-coaches.
For each of the two clubs, the family names and addresses of those born in or after 1981 (children) were then matched with the family names and addresses for those born before 1981 (adults). Where there was a match and the age difference was at least 20 years, it was assumed that there was a child parent grouping. Where there was no match, it was assumed that neither parent of the child was a member of the club.

This list of child only members was then matched by family names and addresses. Where there was a match, it was assumed that the children involved were brothers and or sisters of each other. The list was then compressed by eliminating duplicate family names and addresses. This produced a list of associated parent names and addresses who were not members of either club. This parent list was then sorted against the joining dates of the children involved. Those entries with an earliest joining date of 1998 were discarded. The assumption was that these parents had had less than 1 year's involvement with athletics through their children and would not have been in a position to enter the coach education programme. The remaining family names and addresses were retained for the research. It was assumed that there was one male and one female parent and guardian at each address. Both of these (Associated Parents) became part of the research.

The combined lists of coaches, adult members (non coaches) and associated parents constituted the initial estimate of the research population. The composition of this population estimate by club and by gender is shown in Table 8.2.1.

<table>
<thead>
<tr>
<th></th>
<th>CLUB A</th>
<th>CLUB B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MALE</td>
<td>FEMALE</td>
<td>MALE</td>
</tr>
<tr>
<td>MEMBERS</td>
<td>178</td>
<td>172</td>
<td>203</td>
</tr>
<tr>
<td>PARENTS</td>
<td>93</td>
<td>93</td>
<td>73</td>
</tr>
<tr>
<td>COACHES</td>
<td>15</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>287</td>
<td>277</td>
<td>282</td>
</tr>
</tbody>
</table>

**Table 8.2.1. COMPOSITIONS OF THE INITIAL ESTIMATED RESEARCH POPULATION**
8.3. Sample Size

The maximum sample size for this research was set at 120 for two reasons:

- Adequate time for all the processes in this research only allowed 12 months for interviewing. A longer period of interviewing would have meant that the membership lists used would be significantly out of date. Allowing for holiday periods, this gave about 42 weeks, allowing about 3 interviews per week, a potential of 127 interviews.

- The research anticipated that the logistic regression model with a Goodness of Fit greater than 0.50 would contain between 10 and 15 explanatory variables. For OLS linear regression modelling, these criteria should be achieved with sample sizes significantly less than 120 (Hair, J F, and others, 1995; Odeh, R E, and Fox, M, 1975). For logistic regression modelling, the rules of thumb indicate that a sample size of 120 should be sufficient for between 9 and 17 explanatory variables (Tabachnick, B G, and Fidell, L S, 1997).

8.4. Sampling

The initial number of coaches in both clubs (47) was relatively small, less than 5% of the research population. A 100% sample of the coaches was attempted to capture sufficient information from the large number of variables in the study. It was recognised that some unknown percentage (less than 100%) would yield completed interviews, either because they could not be contacted or they refused to participate. The results of the sample actually obtained have been weighted. The method of weighting is described later in this Section.

To give a maximum sample of 120, a sample of 73 (120-47) was taken from the remaining adults and associated parents in the initial estimate of the research population. Quota sampling was used for these two non coach categories. Each of
these two non coach research populations was divided into 4 groups by club and gender. The quota for each group was proportional to the club/gender proportion within the total non coach population.

The resultant sample quotas for the entire research population (coaches 100%) are shown in Table 8.4.1.

<p>| Table 8.4.1. |
| QUOTA SAMPLES FOR THE RESEARCH POPULATION |</p>
<table>
<thead>
<tr>
<th>CLUB</th>
<th>A</th>
<th>B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEMBERS</td>
<td>13</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>PARENTS</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>COACHES</td>
<td>15</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>30</td>
<td>35</td>
</tr>
</tbody>
</table>

A random quota sample was taken from each of the 8 sub-populations in the first two rows of Table 8.4.1. (members and parents, male and female). Each individual within a sub-population was assigned a random number. Requests for interview were made to individuals in order of their random number. This was repeated until acceptances reached the quota number for each sub-population. A 100% sample was attempted for 4 sub-populations in the third row of Table 8.4.1. (coaches, male and female).

Neither the planned quota samples nor the 100% samples were completely achieved. The 112 interviews obtained are shown in Table 8.4.2.

<p>| Table 8.4.2. |
| COMPOSITIONS OF THE ESTIMATED RESEARCH POPULATION |</p>
<table>
<thead>
<tr>
<th>CLUB</th>
<th>A</th>
<th>B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEMBERS</td>
<td>174</td>
<td>172</td>
<td>199</td>
</tr>
<tr>
<td>PARENTS</td>
<td>93</td>
<td>93</td>
<td>63</td>
</tr>
<tr>
<td>COACHES</td>
<td>19</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>286</td>
<td>276</td>
<td>282</td>
</tr>
</tbody>
</table>

DATA COLLECTION
In the case of coaches, the actual sample differed from the planned 100% for two reasons:

i. Interviews were not obtained from some coaches either because they could not be contacted or because they declined to be interviewed

ii. In interviewing the adult members, 4 additional coaches (all Male, 2 from each club) were encountered who had not been identified either by the club or the SCAA. In all cases they had ceased to be active coaches some years ago. These cases were reallocated from initial estimated research population in Table 8.4.1. to the coach groups from the member groups to give the actual estimated research population shown in Table 8.4.2.

In the case of non coaches, the actual sample differed from the planned quota for three reasons:

i. Interviews were not obtained from some requests either because they could not be contacted or because they declined to be interviewed

ii. Interviews did not reach the quota number because the time allocated for interviewing expired

iii. Interviews exceeded the quota number. Requests for interviews were sent out in batches. All respondents who accepted were interviewed even when the quota number was exceeded. This decision was taken to maintain good will with the clubs.

The resultant actual interview sample obtained and used in the research is shown in Table 8.4.3.
Table 8.4.3.
COMPOSITIONS OF THE VALID INTERVIEWS USED IN THE RESEARCH

<table>
<thead>
<tr>
<th></th>
<th>CLUB</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Members</td>
<td>MALE</td>
<td>FEMALE</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Parents</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Coaches</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>28</td>
<td>25</td>
</tr>
</tbody>
</table>

The response rates achieved and an analysis of the non-responses are discussed in Section 8.8.

Quota sampling produces a weighted random sample, each of the 12 categories in Table 8.4.1. representing a different proportion of the total population. To produce unbiased estimates of the coefficients, the data must be weighted for analysis. How this weighting is constructed and used in SPSS is described in the next Section, 8.5.

8.5. Data Weights & Missing Data

The weighting value is the number of cases in the population represented by one actual sample observation. It is the ratio of the actual sample size to the population size. When weighting is used in logistic regression, the resultant estimates of coefficients remain unbiased. However, the standard errors of the regression coefficients will be underestimated. This drawback is acceptable as the objective of the research is to establish that a logistic model can be made and to determine an order of importance of variables. The objective of the research is not to determine absolute values for the regression coefficients nor to statistically measure the significance of individual coefficients.

SPSS implements data weighting by means of an additional variable 'Weight'. Each case is assigned a value for the variable 'Weight'. This value is the ratio of the number of cases in the population to the number of cases in the sample. The weights used are shown in Tables 8.5.1. and 8.5.2.
Table 8.5.1.
DATA WEIGHTS - ALL 112 CASES USED IN ANALYSIS

<table>
<thead>
<tr>
<th></th>
<th>CLUB</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MALE</td>
<td>FEMALE</td>
<td>MALE</td>
</tr>
<tr>
<td>MEMBERS</td>
<td>14.50</td>
<td>24.57</td>
<td>13.27</td>
</tr>
<tr>
<td>PARENTS</td>
<td>13.29</td>
<td>8.45</td>
<td>7.88</td>
</tr>
<tr>
<td>COACHES</td>
<td>2.11</td>
<td>1.57</td>
<td>1.43</td>
</tr>
</tbody>
</table>

SPSS uses the variable ‘Weight’ to replicate the data by means of simulating the case the number of times given by the variable value. The variable ‘Weight’ can be switched on or off for each analysis. In this research, all analyses were conducted with ‘Weight’ on.

Two of the interviews resulted in two cases having missing values for three explanatory variables. Both cases were male coaches, one from club A and one from club B.

There was no satisfactory method to estimate the missing data values for these two cases. When data from the reduced sample of 110 cases was used, an alternative set of ‘Weights’ were used. The alternative weights are shown in Table 8.5.2.

Table 8.5.2.
DATA WEIGHTS - REDUCED 110 CASES USED IN ANALYSIS

<table>
<thead>
<tr>
<th></th>
<th>CLUB</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MALE</td>
<td>FEMALE</td>
<td>MALE</td>
</tr>
<tr>
<td>MEMBERS</td>
<td>14.50</td>
<td>24.57</td>
<td>13.27</td>
</tr>
<tr>
<td>PARENTS</td>
<td>13.29</td>
<td>8.45</td>
<td>7.88</td>
</tr>
<tr>
<td>COACHES</td>
<td>2.38</td>
<td>1.57</td>
<td>1.54</td>
</tr>
</tbody>
</table>

8.6. Approach to Respondents

The approach, the conduct of the interview and the questionnaire used were developed from 7 pilot interviews conducted during March 1999.
Each individual was sent a letter with a return slip and stamped (2\textsuperscript{nd} class) address reply envelope. An example letter and reply slip are included as Appendices A and B (names and addresses omitted for reasons of confidentiality). There were slight variations according to the club involved and according to whether the individual was a member of the club or the parent of a member.

The salient features of the introductory letter are:

- The letter was on university headed paper to indicate the authenticity of the research
- The reply address was the private address of the researcher in Surrey to indicate local involvement
- The researcher's telephone number and e-mail number were also included to allow alternative contact and a means of checking authenticity
- Reference was made to the appropriate athletic club to confirm that permission had been obtained to use the membership list details
- Names and telephone numbers were given for the chairman of the club, the Manager of the County Council Sports Development Unit and the Head of the University Department to enable checks to be made by the recipient.
- Confidentiality and security were assured.
- In the case of parents, reference was made by name of the children involved
- In the case of parents, an apology was made in case the assumptions about the presence of both parents or the family name were incorrect.
- Only a broad outline of the scope of the research was given with a statement that the research was independent. The objective was to make personal contact with each individual and not to make the research appear too complex or lengthy
- The recipient was given opportunities to refuse or to later withdraw as well as opportunities to control when and where the interview took place.
- The method of follow up was given, but not a specific time interval.
- The incentive was given of contributing personal views to changes in sport.

\textit{DATA COLLECTION}
Response rates, reasons for refusal or non response and methods of response are given and analysed in Section 8.8.

If no reply was received after about 2 weeks, the individual was telephoned at the number shown in the membership list. If this number was found to be incorrect or unobtainable, enquiries were made (including the BT Directory) for a likely alternative. 4 attempts were made to make telephone contact. After this, the case was assigned either to 'No Contact' if no reply was obtained or to a refusal if there was a reply but the individual declined to talk or did not ring back or answer a message left on an answerphone.

Once a positive reply was received, arrangements for interview were made over the telephone. It was made clear that the interview would take about 30 minutes (45 for coaches). Permission was sought to record the interview. Respondents were given full choice of where and when to meet. If asked, the alternatives of at their work, home or athletic club. Confidentiality and security of the data were again stressed. No further details of the purposes of the research were given other than the interest of the researcher into why people were involved with voluntary sport.

Only one objection was made of breach of privacy through use of membership lists. Very few outright and unexplained refusals were received. Surprisingly, given that the researcher is male, just as many men as women requested the interview at their offices or in a public place (the track). No women required that another person be present at the interview. The high rate of success in obtaining interviews indicated that the approach used was both acceptable and practical.

In all cases, the means by which contact was eventually made or attempted was recorded. In the event of a request for an interview failing, including during follow up, as far as possible a reason was sought. Both these sets of information were analysed in order to more fully understand possible relationships of the actual sample.
obtained to the initial estimates of the population. These analyses are reported and
discussed in Section 8.8.

8.7. Conduct of Interviews

The interviews were conducted in three waves:

- September to December 1999. The Adult Members
- January to March 2000. The Coaches

The parents were interviewed first as it was originally thought that they might be
more difficult to achieve the quota samples. This did not prove to be the case (See
Section 8.8.). It was also thought as they were less involved with the clubs, cross talk
between other respondents would be minimised.

The coaches were interviewed last, in the late winter as this is their least busy time.
Also some additional qualitative questions, not directly related to this research, were
added partially as a result of the interviews with parents and members and partially
resulting from developments in the sport.

The adult members were interviewed in the autumn and early winter to avoid holiday
periods and the main competitive season in the summer.

All interviews were conducted by the researcher. Over 90% of interviews were at the
respondent's house. They were mainly during weekday evenings. The remaining
interviews were conducted at the respondents work place or at one of the two club
tracks. In all but two cases, the interviews were conducted with only the researcher
and the respondent present, although children were infrequently present some of the
time.
At the outset of the interview, the consent of the respondent was obtained to record the interview. On no occasion was consent refused. The recording was made in stereo using individual miniature label microphones. Experience had shown that respondents rapidly lost any nervousness with these small microphones and it allowed the seating to be very informal. This proved to be the case. In many instances the respondent finished by saying how much they had enjoyed the interview.

The interview was conducted in accordance with a written questionnaire, an example of which is shown in Appendix C. There were three variations of the questionnaire: for coaches, for members and for parents. During the interview, the researcher filled in the questionnaire and made prompt notes on it. The interview had four sections:

i. A preamble in which a formal statement was read to the respondent specifying the research, the rights of the respondent and the confidentiality of the data. The respondent was then asked to give a name by which he or she would be addressed during the interview to ensure confidentiality if the recording was replayed to other people.

ii. Members were then asked to confirm that they were not and never had been qualified athletics coaches. Coaches were asked to confirm that they were or had been qualified athletics coaches. There were 4 cases, 2 in each club, in which a member respondent stated that he or she had been a qualified coach, but had ceased coaching some years ago. These cases were reallocated to the group ‘Coach’ and the questionnaire changed.

iii. Quantitative questions were asked first, some of which entailed selections from lists or a Likert scale. In these cases, the respondent was handed a show card containing the question and the selection or the Likert scale. They were asked to complete each show card. The show cards are shown in Appendix C at the end of the Questionnaire.
iv. Qualitative questions were asked last. The replies were open ended and very few prompts were used. As far as possible, the qualitative questions were asked as shown in the questionnaire and in the order shown. However, some licence was used if the discussion flowed in a different direction. In some interviews, time precluded all the quantitative questions being fully answered.

Most interviews lasted about 30 minutes. A number lasted more than 1 hour and had to be interrupted to replace the tape. On 5 occasions after about 75-90 minutes, the researcher had to terminate the interview to reach another appointment. On 3 occasions the recording failed. In these three cases, the researcher wrote up the main points of the qualitative elements within 24 hours of the interview from the notes made at the time. All recordings have been retained.

8.8. Response Rates, Reasons & Sample Validity

Table 8.8.1 shows the breakdown of the 204 individuals approached for interview.

<table>
<thead>
<tr>
<th>REQUESTS FOR INTERVIEW</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUB</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>MALE</td>
</tr>
<tr>
<td>MEMBERS</td>
</tr>
<tr>
<td>PARENTS</td>
</tr>
<tr>
<td>COACHES</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Dividing the values in Table 8.8.1. (interview requests) by the values in Table 8.4.3. (valid interviews, the conversion rates achieved have been calculated and are shown in Table 8.8.2. The lower conversion rates for some groups indicate that, in general, more difficulty was encountered in getting interviews with men than with women and with non-coaching members than with coaches and with parents.
Table 8.8.2. 
INTERVIEW REQUEST CONVERSION RATES

<table>
<thead>
<tr>
<th></th>
<th>CLUB</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>Total</td>
</tr>
<tr>
<td>MEMBERS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALE</td>
<td>26.1%</td>
<td>63.6%</td>
<td>45.5%</td>
<td>61.1%</td>
<td>41.7%</td>
</tr>
<tr>
<td>FEMALE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PARENTS</td>
<td>43.8%</td>
<td>78.7%</td>
<td>72.7%</td>
<td>58.3%</td>
<td>62.3%</td>
</tr>
<tr>
<td>COACHES</td>
<td>90.0%</td>
<td>58.3%</td>
<td>72.7%</td>
<td>80.0%</td>
<td>79.1%</td>
</tr>
<tr>
<td>Total</td>
<td>38.9%</td>
<td>67.6%</td>
<td>61.7%</td>
<td>62.9%</td>
<td>54.9%</td>
</tr>
</tbody>
</table>

Lower response rates appear to have been obtained from members, rather than from coaches or parents, and from club A rather than club B.

Before looking at these apparent differences, consideration was given to the mechanisms for case contact and the reasons why interviews could not be obtained. Attempts to contact gave one of three results:

i. Contact initiated by the respondent as a result of:
   - Return of the reply slip in the initial letter
   - Telephone call to the researcher by respondent as a result of the initial letter
   - E-mail set to researcher by respondent as a result of the initial letter
   - Personal contact with the researcher by respondent after the initial letter was sent

ii. Contact resultant from telephone follow-ups by researcher

iii. No contact made:
   - Post Office returned initial letter
   - Unable to contact by telephone, no known number
   - Unable to contact by telephone, X Directory

Overall, 47% of the 204 letters sent out resulted in the respondent contacting the researcher. The predominant means of respondent contact was return of the reply slip.
Contact with 50% of the addressees of the 204 letters sent resulted from follow ups by the researcher. The predominant means was telephone follow ups, usually requiring 2 or more attempts.

The remaining 3% could not be contacted because no telephone number could be obtained or the initial letter was returned marked “Not at this address”.

Analysis of these results showed some differences when categorised by gender and by club:
- More interviews were obtained when contact was initiated by the respondent than when initiated by the researcher.
- Follow ups by the researcher were less successful with men than with women.
- Follow ups by the researcher had to be used more with club A than with club

The full results by gender and by club are shown in Table 8.8.3.

Table 8.8.3.
MEANS OF CONTACT BY GENDER AND BY CLUB

<table>
<thead>
<tr>
<th>INTERVIEW</th>
<th>CONTACT BY</th>
<th>MALE</th>
<th>FEMALE</th>
<th>A</th>
<th>B</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtained</td>
<td>Respondent</td>
<td>51%</td>
<td>57%</td>
<td>49%</td>
<td>58%</td>
<td>54%</td>
</tr>
<tr>
<td></td>
<td>Researcher</td>
<td>49%</td>
<td>43%</td>
<td>51%</td>
<td>42%</td>
<td>47%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Not Obtained</td>
<td>Respondent</td>
<td>34%</td>
<td>52%</td>
<td>38%</td>
<td>42%</td>
<td>39%</td>
</tr>
<tr>
<td></td>
<td>Researcher</td>
<td>71%</td>
<td>40%</td>
<td>71%</td>
<td>47%</td>
<td>55%</td>
</tr>
<tr>
<td></td>
<td>No Contact</td>
<td>5%</td>
<td>8%</td>
<td>1%</td>
<td>11%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>All letters</td>
<td>Respondent</td>
<td>42%</td>
<td>57%</td>
<td>43%</td>
<td>52%</td>
<td>47%</td>
</tr>
<tr>
<td></td>
<td>Researcher</td>
<td>55%</td>
<td>42%</td>
<td>57%</td>
<td>44%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>No Contact</td>
<td>3%</td>
<td>2%</td>
<td>1%</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Whilst no attempt was made to estimate the statistical significance of these figures, some observations may be made.

DATA COLLECTION
The top of Table 8.8.3. shows that, where an interview was obtained, there were few differences in the two means of contact (by the respondent contacting the researcher, by the researcher contacting the respondent).

The middle of the Table indicates some possible differences when an interview was refused (not obtained). A higher percentage women (52%) then men (34%) declined the initial letter. A higher percentage of men (71%) than women (40%) agreed to be interviewed following contact by the researcher. A higher percentage from Club A (71%) than from Club B (47%) agreed to be interviewed following contact by the researcher.

The bottom of the Table shows the response rates where interviews where gained and where they were refused. The two genders and the two clubs show similar patterns.

To look further at the sample, the respondents with whom no interview was obtained were analysed further. No interview was obtained for one of five reasons:

i. Membership Ceased (but not deleted from membership list):
   ii. Resident elsewhere:
       - At university or other education
       - Working away from home
       - Moved away, address unknown

iii. Too busy
iv. Interviewed arranged, but not kept
v. Reason unknown:
   - No known telephone number
   - X Directory
   - No reply to telephone calls

These are analysed in Table 8.8.4.
Table 8.8.4.
REASONS FOR INTERVIEW FAILURE BY GENDER AND BY CLUB

<table>
<thead>
<tr>
<th>REASON</th>
<th>GENDER</th>
<th>CLUB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Membership Ceased</td>
<td>37%</td>
<td>17%</td>
</tr>
<tr>
<td>Resident Elsewhere</td>
<td>19%</td>
<td>28%</td>
</tr>
<tr>
<td>Too Busy</td>
<td>12%</td>
<td>4%</td>
</tr>
<tr>
<td>Interview not kept</td>
<td>1%</td>
<td>8%</td>
</tr>
<tr>
<td>Unknown</td>
<td>31%</td>
<td>44%</td>
</tr>
</tbody>
</table>

The end column of Table 8.8.4. shows that, overall, the predominant reasons were Membership Ceased (29%), Resident Elsewhere, mainly at University, (22%) and Unknown (33%). Analysis by gender showed that cessation of membership was the main reason for no interviews amongst men. Analysis by club showed no marked difference in the pattern of refusal reasons between the two clubs. The analyses by gender and by club are shown in Table 8.8.4.

Further analysis indicated that the rate of cessation of membership was higher in club A than with club B, especially amongst members. This is reflected in the lower response rate from Members of club A.

The conclusions that were drawn from these analyses are:

- The overall response rate of 54.9% (Table 8.8.2.) is probably an underestimate. The actual membership of the clubs is lower than that estimated from the club lists. This is mainly due to the number of names on it of individuals who had, in practise, ceased to be members. This factor was more marked with men than with women; more with members than with parents or with coaches; and more with club A than with club B.

- A proportion of those who could not be contacted over the telephone follow up may have moved and may have ceased membership. Similarly, a proportion of those who are resident elsewhere may have ceased active membership.

DATA COLLECTION
Both these conclusions suggest that the true response rate, the number of interviews obtained divided by the true (active and valid) memberships, is significantly in excess of the 54.9% recorded. This suggests that the response rate achieved is satisfactory, particularly since the proportion refusing or interview not kept was small.

The analysis also suggests that the interview sample achieved is reasonably unbiased in relation to gender and to club. If a bias exists, this may be derived from a biased sample of Members, especially men, due to the greater difficulty in getting interviews with these. Those that did agree to be interviewed may not be completely representative of their sub-populations.

8.9. Sample Validity

Validity is important to anybody trying to replicate research. It is also important in assessing the extent to which the research question has been answered. Two forms of validity are considered:

- Might the sub-populations of interview refusals be non-random samples of their parent sub-populations? If any sub-population is non-random, then the data used in model from that sub-population will introduce a bias. This needs to be considered in interpreting the results.

- Do characteristics of the coaches, members and parents in the two clubs used in the research differ significantly to the same characteristics of athletic clubs in wider contexts of the whole county, Surrey, and nationally? Any differences may be useful in interpreting the results into these wider contexts.

Neither question could be answered as very little comparable data is available. Furthermore, the data that is available may not be very reliable or accurate and may not be entirely comparable due to variations in definitions.
Whilst the objectives of this research are confined to testing a statistical model in a specific decision making in the context of 2 similar athletic clubs, this context can be compared to wider contexts. Such comparisons may assist in interpreting the results to county and national areas. Comparisons can only be indicative as comparable data at county and national levels is very limited and of doubtful accuracy.

The next Chapter describes how the interview variables were selected and how they were incorporated in a questionnaire. The questionnaire is discussed and illustrated. Summaries are given of the sample data collected.
9. THE INTERVIEW & QUESTIONNAIRE

9.1. Introduction

A set of potential explanatory quantitative variables for the logistic regression model was drawn up by reference to existing published research on volunteering and sport, in particular Sundeen, RA. 1992; Lynn, P. 1992; Foster, K et al. 1995; The Sports Council. 1996; Lyle, J et al. 1997; Smith, JD. 1998. These references are collective referred to in this Section, 9.2. as “research related to this Section”. As far as possible, the metric used to measure a variable and the question to elicit the information were those used in previous research. This set of potential explanatory quantitative variables was supplemented by other potential variables resulting from the personal observations of the researcher over 25 years of coaching. This initial set was revised and modified following pilot interviews conducted during the research. Each of the quantitative variables in the revised set is described in Section 9.2.

In addition, other quantitative and qualitative questions were used to provide additional information. The uses of this information are explained below. The details of these questions are given in Sections 10.3. and 10.4.

All interviews were conducted by the one researcher. The interview was designed to produce four sets of data:

i. Quantitative information for direct input into the logistic regression model

ii. Quantitative information to compare characteristics of the sample with other research and, thus, compare the sample with other research sources.

iii. Qualitative information to verify and triangulate the statistical analysis of the quantitative data

iv. Quantitative and qualitative information to aid development in the two clubs and in UK Athletics.
Interviews were constructed to have three parts. The first part had two aims. One aim was to verify that the respondent did belong to the sample group intended (parent, member, coach). The second aim was to ensure that the respondent understood and agreed to the purpose and the conduct of the interview. The second part was to collect the quantitative information chosen for the model and needed to compare the sample with previous research. This quantitative information is reviewed in detail in Sub Section 9.2. of this Chapter.

Both these first two parts of the questionnaire were formally and methodically executed maintaining a strict order of questioning. For these first two parts, the role of the written questionnaire was to ensure consistent presentation.

The third part of the questionnaire was to elicit qualitative reactions to a series of leading questions seeking open ended answers. This part was deliberately administered at the end of the interview. The previous experience of the researcher had shown that respondents replied to such questions much more openly when they had become relaxed during any quantitative part of an interview. This proved to be the case as most respondents replied very willingly and, often, in great depth and length. The role of the written questionnaire in this part of the interview was to give the Interviewer a framework for the entire discussion, rather than a rigid structure.

The qualitative content from the third part of the questionnaire was used for triangulation of the model solution of the quantitative logistic regression model. In the case of an individual who was not a coach, information was sought on the circumstances under which the individual might become a coach. In the case of a respondent who is a coach, information was sought on their present attitudes to and reasons for continuing to coach and how much time they devoted to coaching. In the case of a respondent who was a coach, but was now inactive, information was sought on how long and how frequently they were active and why they ceased. The responses were recorded and later analysed to provide post hoc subjective estimates.
of the probability of a non-coach respondent becoming a coach before the age of 70 and coaches ceasing coaching before the age of 70. The age of 70 was used arbitrarily as an indicator when age related health might inhibit the necessary physical activity required to be a coach.

These triangulation variables are described in Section 9.3. The other information collected to aid the clubs and the sport, but not directly relevant to or used in this research, is reviewed to Section 9.4.

A written questionnaire was used during each interviews to achieve the following:

- To provide prompts to the researcher in conducting the interview to ensure consistency of delivery between respondents in terms of the order of questions; how a question was worded and presented; prompts used; and the major definitions.
- A backup to the tape recording in case of recording failure or poor quality. Also as a means of speeding up the transcription of data from the recording.
- An aid to the researcher verifying information during the interview
- During the qualitative questions, a means of taking notes to prompt the client, especially when a reply was lengthy and went away from the question, and to aid the presentation of subsequent questions. The qualitative section was designed more as a directed conversation than a formal interview.

Three variations of the questionnaire were used, one for each of the sample groups: adult members, parents and coaches. The template was the questionnaire for members. For parents, some of the template questions were amended to refer to the parent and some to refer to the child athlete of the parent. Some template questions not were not applicable to parents and were omitted. The data from these questions was only collected as part of the quantitative and qualitative information for feedback to the clubs. For coaches, most of the template quantitative questions were amended to refer to the year in which the coach entered the coach education programme rather than to the date of the interview. Supplementary questions to the template were
included for coaches to allow comparisons with previous research and for guidance for future club and sport development.

One effect of having three variations of the questionnaire was that a set of all three variations had to be taken to every interview in case the respondent turned out to be in a sample group not as originally understood. This occurred four times when a member was found to be a qualified coach.

The questionnaire was not given to the respondent and the completed questionnaire was not shown to the respondent. However, five quantitative questions required choices to be made by the respondent. In these cases, a printed Show Card was given to the respondent containing the question with either a selection of answers or a Likert scale. The respondent was provided with a pen to mark their responses. Each Show Card was handed to the respondent at the time that that question was asked. The completed Show Cards were collected and are stored with the corresponding completed questionnaire.

A questionnaire and a set of Showcards for members are shown in Appendix C. The names of the clubs have been omitted from this Appendix to preserve their anonymity. The order of the choices on the Showcards was random and four different sets were randomly used to minimise any effect of order on respondent choice.

The questionnaire and the interview methodology were the result of modifications from an initial format that was piloted 3 months before the interviews began using 6 respondents: 3 from each club and a member, a parent and a coach from each club. Both male and female respondents were chosen for the pilot interviews. The modifications made were ones of detail, wording and sequence rather than fundamental changes.

The response variable, $Y$, in this research is referred to as “Coach Decision”. $Y$ is coded as “1” if the respondent held or had held a valid coaching award at Level 1 (Assistant Club Coach) or above made by the NGB of athletics in the UK. Otherwise, $Y$ is coded as “0”. Verification of statements by respondents was made from Coaching Lists supplied by Surrey County Athletic Association or from the chairman of the club. Awards can only be obtained by entering the Coach Education Programme.

The underlying hypothesis is that the probability of an individual entering the UK athletic Coach Education Programme is significantly associated with a set of explanatory variables. The corresponding null hypothesis is that Goodness of Fit measures for a logistic regression model of the response variable, $Y$, using a set of explanatory variables, $X$, exceeds 0.90 or 90.0%. The Goodness of Fit measures used are discussed in Chapter 10.

An equivalent Confidence Level of 0.01 for significance was used for the fitting of the logistic regression model. This level of confidence is considered to be adequate for behavioural studies of this nature (Sundeen, R A, 1992).

The year in which an individual entered the Coach Education Programme and the highest level of award obtained were recorded, together with the last year in which the individual coached at least once a week, the current frequency of coaching, the amount of time given to coaching and any reasons for ceasing coaching (where applicable). This information was collected for comparative purposes and for feedback to clubs and the NGB, not for the logistic regression.
9.2.1. GENDER

Much research on volunteering and sport has noted significant differences between the genders in terms of probability of volunteering, of involvement in sport, and in becoming a coach (Anderson, J C, and Moore, L F, 1978; Humble, S, 1981; Jackson, H, 1985; Williams, J A, and Ortega, S T, 1986; Willis, P L, 1998). The explanatory variable $X_1$ was coded as “1” if Male, otherwise “0”. The hypothesis is that males are more likely to become coaches than females. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with gender $X_1$. This is indicated when the log likelihood associated with the addition of $X_1$ to the logistic regression model is not significant. This measure of association was used to test all other explanatory variables.

In the context of a logistic regression model (using explanatory variables, each with a unique coefficient, which are additive in effect) the above hypothesis assumes that the same set of significant explanatory variables are associated with female respondents as are associated with male respondents, i.e. that the entire population of male and female respondents is homogeneous with respect to a single significant set of explanatory variables. The assumption also implies that the direction of the association of any explanatory variable is the same for men as for women.

For example, men are more likely to be in full time work that involves commuting. The time at which they are free from work and commuting and can be at a track during the week may be crucial in their decisions regarding their leisure. On the other hand, women are more likely to be in part time work or occupied full time with household tasks. The time at which they can be free for leisure may be more flexible than for men. Consequently, the time at which a woman considers herself free may be less important in her decision making than for men.

This assumption of homogeneity will be examined in testing the model.

INTERVIEW & QUESTIONNAIRE
9.2.2. DEPENDANTS

Dependants were defined as all natural and adopted and step children plus spouse or partner plus any other individuals of any age who are dependent upon the respondent for care through age, health, injury or other impediment and are living permanently in the same house. Numbers of dependants were collected in 8 categories:

- Males under the age of 18, not associated with athletics, \( X_2 \) referred to as “U18Mnot”;
- Females under the age of 18, not associated with athletics, \( X_3 \) referred to as “U18Fnot”;
- Males under the age of 18, associated with athletics other than at school, \( X_4 \) referred to as “U18Myes”;
- Females under the age of 18, associated with athletics other than at school, \( X_5 \) referred to as “U18Fyes”;
- Males under the age of 18 and over, not associated with athletics, \( X_6 \) referred to as “Mnot”;
- Females under the age of 18 and over, not associated with athletics, \( X_7 \) referred to as “Fnot”;
- Males under the age of 18 and over, associated with athletics other than at school, \( X_8 \) referred to as “Myes”;
- Females under the age of 18 and over, associated with athletics other than at school, \( X_9 \) referred to as “Fyes”.

The published research related to this Section on the influences of dependants on volunteering and coaching is not conclusive. This may be due to variations in the collection of evidence. The measurements of dependants vary across counting only children, all members of household, only those involved in the activity of respondent, etc. Studies of volunteering indicate that numbers of dependants can be either negatively or positively associated with the probability of volunteering or not significant. Studies of sports volunteers, especially athletics, indicate that involvement of children and partners is a major reason for parents to volunteer, especially as

INTERVIEW & QUESTIONNAIRE
coaches. There is some anecdotal evidence that parents of girls are more likely to coach than parents of boys. The age of 18 is used to define a child as this was the minimum age for entry into the Coach Education programme.

An issue addressed by this variable concerns the proportion of the dependants that are involved with athletics. The concept is that the lower this proportion is, the greater the pressure on the respondent to minimise time involved in athletics. For example, a mother with children not involved with athletics will have less motivation to become involved with athletics than if all her children were involved in athletics. This may be because the children are too young or because they have other interests. Similarly, a single parent with only one of several children involved will have less motivation than a two parent family.

This data can be used in conjunction with another variable, transport (see Section 9.2.4), to indicate transport conflicts within a household both between adults in the household and between individuals, especially children, with interests other than athletics.

The hypotheses associated with these variables are that the fewer the number of dependants associated with athletics, the lower the probability of the individual deciding to become a coach. The variable may interact with another explanatory variable, cars available. The Null Hypotheses are that the probability of becoming a coach \( Y^* = 1 \) is not significantly associated with variables \( X_2 - X_9 \).

9.2.3. DISTANCE

Logically, the distance from home to the athletics venue may be a consideration when an individual perceives that they have a shortage of free time or there are conflicts on how the free time is used or on the cost of transport. Research related to this Section on volunteering frequently presents quotations such “Do not have sufficient time” as a reason for not volunteering or for stopping voluntary activity.

INTERVIEW & QUESTIONNAIRE
The address of the respondent was confirmed. The OS map grid co-ordinates of the address were found, together with those of the athletics club they used. The linear, crow-fly, distance was calculated between the address and the club. Explanatory variable $X_{10}$, “Dist” was given this value.

Respondents were also asked to estimate how long, in minutes, it took them to reach their track from their home. The estimate was their perception of duration, rather than the actual time. The explanatory variable “Time” was given this value and used as an alternative to the variable “Dist”.

It is recognised that there may be an interaction between these variables and another potential variable, Free Time, described later in Sub-Section 9.2.6.

The hypothesis for $X_{10}$, “Dist”, is that the nearer an individual lives from the track, the higher the probability of the individual deciding to become a coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with $X_{10}$.

The alternative hypothesis for $X_{10}$, “Time”, is that the less time an individual perceives it takes to get to the track, the higher the probability of the individual deciding to become a coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with $X_{10}$.

9.2.4. TRANSPORT

Research related to this Section on volunteering suggests that people may be deterred from volunteering if they do not have access in their free time to a car. The decline of public transport, especially in the evening, and the perceived risks to children from abuse, now make lack of personal use of a car a deterrent to any participation in sport for both adults and children, whether or not coaching is involved.
Research related to this Section on volunteering has consistently indicated that the rate of volunteering is positively associated with income and socio-economic grouping. However, the researcher felt that questions on income might be embarrassing and could jeopardise complete interviews. In addition, data on both income and on socio-economic group are very subject to problems of definition. As a result, it was decided to use the number of cars or vans available after work as a surrogate for estimates of income or socio-economic group. It was also recognised that this variable might interact with explanatory variables $X_2\ldots X_9$, see Section 9.2.2.

Respondents were asked to state how many cars and or vans were normally available to them outside work hours.

The hypothesis is that the more cars/vans an individual has access to after work, the higher the probability of the individual deciding to volunteer as a coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with $X_{11}$.

9.2.5. EDUCATION

Research related to this Section has looked at either the age at which full time education ceased or the highest level of educational attainment. Some of this research has indicated that these both appear to be positively associated with becoming a volunteer. But a survey of Scottish survey found that coaches in athletics had significantly fewer years of formal education than coaches in other sports (Lyle, J, Allison, M, and Taylor, J, 1997). What this survey did not measure was the average characteristics for people associated with all the sports surveyed or for any of the individual sports, including athletics. From this research, no conclusion can be drawn about any association of education with becoming a coach.
In all the research, there are problems with the instruments. Age at which education ceased usually fails to take into account either of part time education following leaving full time school or college or of education resumed later in working life or during periods of unemployment or on retirement. Educational level attained is a categorical variable in which definitions of the categories are often ambiguous, not fully inclusive and varying with time and country. Categories are often presented as ranked, but without stating the criteria for ranking other than tradition, e.g. a degree is ranked above a vocational college award that is ranked above school leaving awards.

In this research, two separate instruments have been used:

i. The number of years of full time school education from the age of 15, $X_{12}$ referred to as “School”.

ii. The number of years and fractions of a year of full or part time education since leaving school, $X_{13}$ referred to as “FEHE”.

The age of 15 was chosen as that is likely to be the lowest statutory school leaving age. This age has been in force in the UK covering the age span of the populations sampled in this research.

Each respondent gave their year of birth and the year in which they considered that they left school. These allowed Item (i) above, number of years of full time school education from the age of 15, to be calculated.

Each respondent stated the amounts of full or part time education that they perceived that they had had since leaving school. In the case of full semester education, including university and colleges of further education, the number of complete years were given. Details were also asked of any part time education or formal courses. These were defined as formal taught courses, academic and vocational, of at least two weeks duration, but less than a semester including day release, apprenticeship, part
time and evening courses. Where a course was full time, the respondent was asked to state the number of weeks duration. Where a course was part time, the respondent was asked the number of days or hours per week and the duration of the course. 1 day per week was measured as 0.2 years per year. 1 evening per week was measured as 0.1 years per year. The total gave Item (ii), number of years and fractions of a year of full or part time education from leaving school.

An alternative potential variable, “Educat”, was available as the summation of “School” and “FEHE”.

The hypotheses for all three potential variables are that the greater the number of years of education, the higher the probability of the individual deciding to volunteer as a coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with years of formal education.

9.2.6. FREE TIME

Research has indicated that athletic coaches give an average of 12 hours a week (Lyle, J, and others, 1997, page 26, Table 10). Other research indicates that the time given by sports volunteers has a highly skewed distribution with a high variance (The Sports Council, 1996 (a)). As the amount of time required to coach is considerable, the amount of free time available of an individual can be expected to be a significant factor in deciding to coach, especially because much of the time will be after work during weekdays. Whilst formal working hours declined during the 20th Century, recent work trends appear to have increased stress of work and, hence, the amount of free time needed to recover. Other trends effecting free time are increases in the amount of work which individuals now have to complete at home (partially related to the appearance of portable and home computers and the Internet) and the increase in both partners having full or part time jobs.
The last few decades have also seen social changes that tend to increase the demands on the free time of parents. One is the demise of public transport. Allied to this is the enormous increase in the availability of cars for private use. At the same time there has been an increase in the perception of risk of violence or abuse in public places. All these have resulted in parents having to spend more and more of their free time in transporting children by car.

Each respondent was asked to state the earliest time on a Tuesday or Thursday evenings on which the respondent was free from business, education and commuting or normal household duties. In the case of shift workers, activities include any necessary sleeping and resting during daytime. In case of respondents engaged on work taking them away from home, e.g. airline staff, the hour related to periods when they are not working away from home. Tuesdays and Thursdays were chosen as these are the normal weekday evenings that local athletics clubs and the coaches operate. The results were recorded as $X_{14}$, "Free", as the hours and decimal fraction of the hour after 12 noon, e.g. 4.15 p.m. or a quarter past four in the afternoon recorded as $X_{14} = 4.25$.

Respondents were also asked if, during the past 12 months, their work involved them in shifts or irregular hours (including overtime) or frequently (at least twice a month) staying away from home. The potential explanatory variable, $X_{15}$, "Irregular", was coded "1" if the respondent responded positively to any of these three causes of irregular working, otherwise $X_{15}$ was coded "0".

The hypothesis related to potential explanatory variable $X_{14}$ was that there was a positive association between the amount of free time (or a negative association with the hour at which free time commenced) with the probability of the individual deciding to volunteer as a coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with the amount of free time (or the hour at which free time commences.
The hypothesis related to potential explanatory variable $X_{15}$ was that working irregular hours decreased the probability of the individual deciding to volunteer as a coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with irregular hours.

9.2.7. ASSOCIATION WITH SPORT

Volunteering was long considered to have a high degree of altruism: serving others without regard for personal benefit. As volunteering became a subject for serious research, it was realised that voluntary organisations could be seen as two groups, each with differing motivations:

i. Motivated by organisational and or individual objectives, primarily related to the provision of financial help and services to others. Volunteering takes place to help others without expectation of tangible reward for the volunteer or for the volunteer’s immediate family. This is the traditional view of altruistic volunteering. Many of these organisations form the Registered Charities.

ii. Motivated by desires for mutual self help amongst their members and their families, including the individual. Volunteering takes place to benefit oneself and, possibly, one’s family from a collective interest and is to help both others and oneself to attain shared aims.

Most amateur sports clubs fall into the latter category, the aims being to share expensive facilities and to organise competition. Research related to this Section has indicated that members of sports organisations usually join for personal benefit or for benefit to their immediate family. With an increasing adult public interest in relations between exercise and health, the benefit sought is often that of personal health. Such motivations are particularly strong where the sport primarily concerns individual activity rather than that of a team. Athletics is one of these sports.
However, sport is increasingly becoming the basis of a career. Participation in an amateur club and of having coaching qualifications is increasingly being used to gain professional recognition on CV’s and to gain knowledge to be used in a paid job. Initially, such participation was confined to teachers and a very limited number of local authority posts involved with leisure. The growths of both the leisure industry (especially leisure centres and health clubs) and leisure retailing (especially clothing and footwear) have both added to paid jobs directly related to sport. Growing interest in the treatment of sports injury has lead to increases in the number of people earning a living through this field. They, too, perceive participation in a local sports club as being beneficial to their career.

The sharp increase in the number of professional players and coaches (full and part time) of sport, including those receiving sponsorships and financial help from public and charitable funds, such as the National Lottery, has produced other personal, non altruistic, motivations to participate.

The recent growths over the past two decades of veteran sport and of mass participation Marathons have both extended the age range of personally competing in and continuing active association with sport. Both these activities tend to take up a significant amount of the free time of the individual because of the amounts of training required. One result may be that these forms of participation in sport may substitute for becoming a coach. This may be particularly true of mass participation running as it gives great freedom in scheduling personal time as the facilities, including coaching, needed are limited. The motivation and attraction may be to retain association primarily for personal benefit with minimal responsibility and maximal flexibility of time.

In the case of parents, association with sport and the related possibility of volunteering as a coach may be more through a child than for the individual. The motivation is often not at all altruistic and solely to ensure that the child or children receive maximum benefit from a sports club. Such associations may, over time,
change to a motivation of helping others, not just family. However, even that
motivation is not strictly altruistic as the individual usually gains personal benefit from
enjoying the activity and being part of an organised society.

Thus, association with sport and how it may or may not develop into a desire to
volunteer as a coach is a complex set of reactions to circumstances and of
motivations. A number of measures of association and of motivation have been used
in this research, although no claim is made by the researcher that all aspects have been
covered. The following potential Explanatory variables were measured through the
interviews:

i. Whether or not members of the family were involved in any other sports,
other than athletics outside school, $X_{16}$, referred to as “Other”. $X_{16}$ was
set at “1” if there was any other involvement, a positive answer, otherwise
set at “0”.

ii. The percentage of years since leaving school that was spent in regularly (at
least once a season) competing in any sport, including athletics, including
years when injured or abroad, $X_{17}$, referred to as “Compete”. The
respondent was asked to estimate the number of years they had been
involved in any competitive sport since they left school. The number of
years since the respondent left school was calculated and the estimated
number of years of competitive sport was expressed as a percentage of
this.

iii. Whether or not the individual has ever competed or intended to compete
as a veteran in athletics, $X_{18}$, referred to as “Vet”. $X_{18}$ was set at “1” if
there was a positive answer, otherwise set at “0”.

iv. Whether or not the individuals main interest in athletics was road running,
including mass participation events and marathons, $X_{19}$, referred to as
“Road”. $X_{19}$ was set at “1” if there was a positive answer, otherwise set at
“0”.

v. Using a 5 point Likert Scale, each respondent indicated the frequency over
the past 12 months with which the individual was helping with or

**INTERVIEW & QUESTIONNAIRE**
competing in any formal organised sport or exercise (other than recreational walking), \(X_{20}\), referred to as “Part”. The scale ran from “Once a week or more” (1), through “Fairly regularly, but not every week” (2) and “Irregularly, at least once a month” (3) and “Very infrequently, not usually once every month” (4) to “Not at all” (5).

vi. The respondent was asked to estimate, over the past 12 months, the percentage of work or study time involved with or related to sport, \(X_{21}\), referred to as “Work”.

The hypothesis related to potential explanatory variable \(X_{16}\), “Other”, was that having family interests in one or more sports other than athletics decreased the probability of the individual deciding to volunteer as an athletics coach. The null hypothesis is that the probability of becoming a coach \((Y^* = 1)\) is not significantly associated with having family interests in sports other than athletics.

The hypothesis related to potential explanatory variable \(X_{17}\), “Compete”, was that the higher the percentage of post school years that were spent in competing in sport, the higher the probability of the individual deciding to volunteer as a coach. The null hypothesis is that the probability of becoming a coach \((Y^* = 1)\) is not significantly associated with the percentage of post school years spent in competing in sport.

The hypothesis related to potential explanatory variable \(X_{18}\), “Vet”, was that being or intending to be a veteran athlete decreased the probability of the individual deciding to volunteer as a coach. The null hypothesis is that the probability of becoming a coach \((Y^* = 1)\) is not significantly associated with being or intending to be a veteran athlete.

The hypothesis related to potential explanatory variable \(X_{19}\), “Road”, was that having prime interest in athletics in road running decreased the probability of the individual deciding to volunteer as a coach. The null hypothesis is that the probability of becoming a coach \((Y^* = 1)\) is not significantly associated with having a prime interest in athletics in road running.
The hypothesis related to potential explanatory variable $X_{20}$, "Part", was that the higher the current amount of participation (of any kind) in sport, the higher the probability of the individual deciding to volunteer as a coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with the amount of current participation in sport.

The hypothesis related to potential explanatory variable $X_{21}$, "Work", was that the higher the percentage of work involving sport, the higher the probability of the individual deciding to volunteer as a coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with the percentage of work involving sport.

9.2.8. LITERACY

Studies have repeatedly shown associations between educational age and reading ability (Gilliland, J, 1972; Davison, A, and Green, G M, (Eds.), 1988; Singar, H, and Donlan, D, 1989). A Sports Council survey (The Sports Council, 1996 (a)) of coaches in 1996 showed that 48% of all coaches had an educational achievement level of below ‘A’ Levels. A 1997 survey (Lyle, J, Allison, M, and Taylor, J, 1997) of sports coaches in Scotland showed that 28% of all coaches had left school by the age of 19. However, for athletics coaches, a significantly higher percentage, 65%, had left school by 19.

This research obtained data (Surrey County Athletics Association, 1998; Central Statistical Office; 1992) allowing a comparison between the density of athletics coaches in the populations of the 15 wards of the Historic County of Surrey with the percentage of low literacy found in each ward (Basic Skills Agency, 1998). The Pearson Correlation Coefficient of -0.63 ($\alpha = 0.01$; $N = 15$) was significant, indicating that there was an association between literacy level and the relative number of athletics coaches in a population.
These studies indicate that reading ability may be an important factor in deciding to become a coach. It must be expected that educational level may be associated with reading ability. Consideration was given to applying accepted literacy tests to the interview sample. This was rejected on the basis that such tests are time consuming, need specialist training and can be intimidating. Instead, it was decided to measure the readability of the normal leisure reading of the interview respondents. Readability is the study of the match between a reader and a text Gilliland (1972) citing Dale, E and Chall, J S (1948). The ease with which an individual can read a given text is partly conditioned by the motivation of the individual (Klare, G R, 1963; Gilliland, J, 1972; Harrison, C, 1980; Davison, A, and Green, G M, (Eds.), 1988). It was for this reason of motivation that a decision was taken to measure readability from the newspaper or magazine most frequently read by each respondent, these being their preferred leisure reading material (Nicoll, A, and Harrison, C, 1984).

The metric chosen for readability was the well established Flesch Ease of Reading Score (Flesch, R F, 1948; Lunzer, E, and Gardener, K, 1979). It is one of the few measures of readability developed for adults rather than for children. Flesch found that word length correlated (r = 0.87) with word complexity, an indirect measure of abstraction. He also found that sentence length correlated (r = 0.72) with sentence complexity. The metric is constructed so that most scores lie between 0 and 100, although it is possible for extreme scores to lie outside this range. It should be noted that the HIGHER the score, the EASIER, the reading. Texts which are difficult to read will have low scores. Texts which are easy to read will have high scores. Equation 9.2.8.1. gives the formula for calculating a Flesch Score:

\[ RE = 206.8 - 0.8 \cdot w1 - 1.0 \cdot s1 \]  

(Equation 9.8.2.1.)

where:

- **RE** is the Flesch Ease of Reading Score
- **w1** is the measure of word length, in terms of syllables per 100 words.
- **s1** is the measure of sentence length, in terms of words per sentence.
As the number syllables per word and the number of words per sentence have to be measured, these two variables, being indirect measures of complexity and abstraction, were used as potential Explanatory variables in addition to the Flesch Score.

Flesch suggested that scores should be calculated from two or three samples of 100 consecutive words from a text, the samples being taken from the beginning, middle and end of the text. This sample size was probably chosen for convenience as computers were not available in the late 1940’s when the metric was first developed. No statistical evidence could be found showing what might constitute a statistically significant and safe sample size or sampling procedure.

A number of computer programmes exist which purport to measure Flesch Scores, the most readily available being in MS Word. A study for The British Library (Sydes, M, and Hartley, J, 1997) concluded that this programme was unreliable. This research undertook further trials, comparing the computer results with manual counts, and came to the same conclusion.

This research involved studies to establish a statistically reliable sampling procedure and sample size. A semi-computerised method was developed for computing syllable and word counts and Flesch scores from newspaper and magazines texts. This uses MS Excel. These studies found that:

- the minimum reliable sample size from newspapers was not less than 500 words in not less than 20 sentences randomly chosen from an article.
- there are significant differences in Flesch Scores between newspapers.
- there are significant differences within a newspaper between articles mainly written about people and containing quotations (a “People” perspective) and those mainly written about facts (a “Fact” perspective). Generally, “People” articles have higher Flesch Scores (easier to read) than “Fact” articles.

INTERVIEW & QUESTIONNAIRE
• within a newspaper, a perspective and a writer, differences in Flesch Scores between the subjects of articles (sport, arts, politics, etc) were not highly significant.

• within a newspaper, a subject and a perspective, differences in Flesch Scores between writers were not significant.

As a result, Flesch Scores were computed for every newspaper mentioned by respondents and for both perspectives within each newspaper. The two samples from each newspaper were chosen at random from all the articles in that newspaper.

Diagram 9.2.8.1. illustrates the ranges of Flesch Scores encountered and the variations between newspapers and between the two perspectives. It will be seen that the broadsheet newspapers tend to have lower Flesch Scores than the tabloids, corresponding to the ease with which they can be read and the literacy level required. It will also be seen that, for any given newspaper, “People” articles tend to be easier reading than “Fact” articles, although the difference is less with the tabloids.
The full results are shown in Table 9.2.8.1.

Table 9.2.8.1.
LITERACY MEASURES - BY PUBLICATION AND PERSPECTIVE
(Ordered by Flesch Score, hardest reading first)

<table>
<thead>
<tr>
<th>PUBLICATION</th>
<th>FACTS</th>
<th>PEOPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Flesch Score</td>
<td>Syllables per Word</td>
</tr>
<tr>
<td>Financial Times</td>
<td>42.05</td>
<td>1.63</td>
</tr>
<tr>
<td>The Independent</td>
<td>46.68</td>
<td>1.61</td>
</tr>
<tr>
<td>Horse &amp; Hounds</td>
<td>46.93</td>
<td>1.52</td>
</tr>
<tr>
<td>The Times</td>
<td>48.87</td>
<td>1.59</td>
</tr>
<tr>
<td>The Guardian</td>
<td>49.50</td>
<td>1.62</td>
</tr>
<tr>
<td>Evening Standard</td>
<td>49.65</td>
<td>1.57</td>
</tr>
<tr>
<td>Daily Telegraph</td>
<td>50.02</td>
<td>1.56</td>
</tr>
<tr>
<td>(Local) Review</td>
<td>50.58</td>
<td>1.55</td>
</tr>
<tr>
<td>(Local) News &amp; Mail</td>
<td>51.21</td>
<td>1.54</td>
</tr>
<tr>
<td>The Observer</td>
<td>51.38</td>
<td>1.56</td>
</tr>
<tr>
<td>Surrey Advertiser</td>
<td>52.62</td>
<td>1.57</td>
</tr>
<tr>
<td>Racing Post</td>
<td>54.39</td>
<td>1.46</td>
</tr>
<tr>
<td>The Express</td>
<td>55.73</td>
<td>1.55</td>
</tr>
<tr>
<td>Cosmopolitan</td>
<td>56.75</td>
<td>1.57</td>
</tr>
<tr>
<td>The Daily Mail</td>
<td>57.18</td>
<td>1.53</td>
</tr>
<tr>
<td>The Daily Star</td>
<td>60.65</td>
<td>1.49</td>
</tr>
<tr>
<td>The Mirror</td>
<td>62.61</td>
<td>1.53</td>
</tr>
<tr>
<td>Women's Weekly</td>
<td>64.34</td>
<td>1.46</td>
</tr>
<tr>
<td>The Sun</td>
<td>64.54</td>
<td>1.48</td>
</tr>
<tr>
<td>News of the World</td>
<td>69.08</td>
<td>1.41</td>
</tr>
</tbody>
</table>

n/a = Not Available.

This shows that whilst Word Complexity (Syllables per Word) generally follows the rank order of the Flesch Scores, there is considerable variation from this rank order for Sentence Length (Words per Sentence), for example The Racing Post.

Each respondent was asked to name the newspaper most frequently read and which of the two perspectives were read (or both). From these responses, the appropriate values could be allocated from Table 9.2.8.1. for each respondent to Flesch Score, X_{22}, "Flesch"; Syllables per word (word complexity), X_{23}, "Word"; and Words per sentence, (sentence length), X_{24}. Where a respondent stated that they read both perspectives, the lower value of Flesch Score was allocated (hardest reading), together with the corresponding values for Syllables per word and Words per sentence.

Each respondent was asked to indicate which of 10 subjects in their newspaper that they read regularly. One category was science and technology. The 10 subjects were
presented to each respondent on a Show Card (see Appendix C). The order of subjects on the Show Card was random. 4 Show Cards were randomly used, each with a different subject order.

The value of $X_{25}$, "Science", was set at "1" if "Science and Technology" was marked on the Showcard, otherwise $X_{25}$ was set at "0". This variable was included in case lack of familiarity with the subject might be a deterrent to respondents choosing to become a coach.

Because they are related and are likely to be highly correlated, in constructing the model, explanatory variables $X_{23}$ and $X_{24}$ (word complexity and sentence length) or explanatory variable $X_{25}$ (read science) were used as alternatives to explanatory variable $X_{22}$ (Flesch Ease of Reading Score).

The hypothesis related to potential explanatory variable $X_{22}$, "Flesch", was that having a high value of Flesch Ease of Reading Score of leisure reading literacy decreased the probability of the individual deciding to volunteer as an athletics coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with leisure literacy, as measured by the Flesch Ease of Reading Score.

The hypothesis related to potential explanatory variable $X_{23}$, "Word Complexity", was that having a low value of syllables per word of leisure reading decreased the probability of the individual deciding to volunteer as an athletics coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with leisure literacy, as measured by syllables per word.

The hypothesis related to potential explanatory variable $X_{24}$, "Sentence Length", was that having a low value of words per sentence of leisure reading decreased the probability of the individual deciding to volunteer as an athletics coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with leisure literacy, as measured by words per sentence.
The hypothesis related to potential explanatory variable $X_{25}$, “Science”, was that reading articles on science and technology for leisure increased the probability of the individual deciding to volunteer as an athletics coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with reading science and technology articles for leisure.

9.2.9. ATTITUDE TO VOLUNTEERING

A US Gallop survey on volunteering using logistic regression, three attitudinal explanatory variables were included. Whilst none of these were found to be significant for recreational organisations, one was found to be significant for nonprofit, non-sectarian organisations. The other two attitudinal variables were again found to be not significant (Sundeen, R A, 1992). The significant variable related to the attitude of respondents to whether or not people should volunteer. The instrument designed for this variable was used in this research to give values for explanatory variable $X_{27}$, “Vol”.

Each respondent was given a Showcard with a question and a 5 point Likert scale to indicate their response to the question. The question posed was that “People should volunteer some of their time to people elsewhere”. The 5 points of the scale ran from Strongly Disagree (value 1), through Disagree and Neither (disagree or agree; or Don’t Know) and Agree to Strongly Agree (value 5).

The hypothesis related to potential explanatory variable $X_{27}$, “Vol”, was that the stronger the attitude to the statement that people should volunteer, the higher the probability of the individual deciding to volunteer as an athletics coach. The null hypothesis is that the probability of becoming a coach ($Y^* = 1$) is not significantly associated with the attitude that people should volunteer some of their time to people elsewhere.
9.2.10. ATTITUDE TO PAYING COACHES

The rise of local authority leisure centres in the 1970’s, followed by the rapid increase in profit making health clubs in the 1980’s, both reported in Chapter 1, led to a marked increase in paid posts in all aspects of sport and leisure. As reported in Chapter 3, the 1990’s saw most sports admit professional athletes, albeit still a small proportion of the total number of participants in sport, and the conversion of some of the larger clubs, especially the major football clubs, into profit seeking quoted companies. Along with professional players came professional administrators, managers and coaches. The trend to professional administration and coaching has been boosted by government policy aimed at improving the international performances of UK sport and using Lottery money to create and support several hundred posts of sports development managers. All this may have created an expectation amongst the general public that quality coaching in sport requires payment. This would be in line with the general rise in the concept of the consumer society.

This trend and expectation of paid coaching has to be set against the fact that the majority of coaches in athletics in the UK are unpaid (The Sports Council, 1996(a)) and that the tradition of amateurism prevails at club level. In preparing for this research, the researcher found an increasing resentment at club level to these increases in paid posts in sport. Whilst some of this may stem from envy, much of it appears to stem from a genuine desire for sport at local level to remain amateur and unpaid. The concept appears to be that payment alters the motivation of individual away from serving the local community with minimal regard to personal benefit, the original basis of volunteering and charity, to preserving and improving a personal career. There is a genuine fear that the employment of even a few coaches would cause tensions with the unpaid volunteers and that the power and influence of these paid coaches might dominate the scene.
It might be expected that people who volunteer to be unpaid coaches may hold strong views about payment of coaches. At the same time, it is recognised that such views may be generational and may not be held so strongly by members of more recent generations who have only experienced the consumer society. The generational aspect is discussed more in Sub-Section 9.2.11, which follows and is recognised that this generational variable may interact with the payment attitudinal variable discussed in this Section, 9.2.10.

A literature search failed to find an existing and proven metric for measuring attitude to payment of coaches. Instead, the 5 point Likert scale used in the previous Section, 9.2.9, to measure attitudes to volunteering was adapted for the purpose.

Each respondent was given a Showcard with a question and a 5 point Likert scale to indicate their response to the question. The question asked the respondent to indicate the extent to which they agreed or disagreed with the statement that “coaches at local sports clubs, generally, should be paid”. The 5 points of the scale ran from Strongly Disagree (value 1), through Disagree and Neither (disagree or agree; or Don’t Know) and Agree to Strongly Agree (value 5).

The hypothesis related to potential explanatory variable X_{27}, “Pay”, was that the stronger the attitude to the statement that coaches should be paid, the lower the probability of the individual deciding to volunteer as an athletics coach. The null hypothesis is that the probability of becoming a coach (Y^* = 1) is not significantly associated with the attitude that coaches should be paid.

9.2.11. AGE

Research related to this Section indicates that the probability of volunteering and of becoming a coach increases with age to a certain point and then decreases, especially around the retirement age of 60. Research specific to sport report that that the peak
age is approximately 40 years old for athletic coaches (The Sports Council, 1996 (a); Lyle, J, and others, 1997).

The research does not show the causation of peak age. Rapoport, R & Rapoport, R.N. (1975) hypothesised that there were Events in the Family Life Cycle that tended to be associated with individuals making decisions on volunteering. Two such Events frequently occur at about the age of 40:

- Children reaching the age of about 10, the age at which many start sport in clubs outside school.
- Reaching stability and peak earnings in career.

Also from about this age, accelerating during the 50’s and 60’s, physical powers begin to reduce and often result in the individual taking less interest in participating in sport. There is some overlap with this age range with another Family Life Cycle Event, retirement. Some of the research indicates that individuals nearing or into retirement do not volunteer unless they had experience of volunteering before retirement. Rapoport, R & Rapoport, R.N. 1975 consider retirement to be another Event in the Family Life Cycle.

It can be hypothesised that the peaking and subsequent decline in becoming a coach with age may be associated with Life Cycle Events. This makes the relationship between age and volunteering to be a coach non-linear. To allow for this, the continuous variable “age” has been transformed into a categorical explanatory variable, as shown in Table 9.2.1.

<table>
<thead>
<tr>
<th>Category Coding</th>
<th>Age Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18-24</td>
</tr>
<tr>
<td>2</td>
<td>25-34</td>
</tr>
<tr>
<td>3</td>
<td>35-44</td>
</tr>
<tr>
<td>4</td>
<td>45-54</td>
</tr>
<tr>
<td>5</td>
<td>55-64</td>
</tr>
<tr>
<td>6</td>
<td>65+</td>
</tr>
</tbody>
</table>
Variable $X_{28}$, “Age”, was allocated the category coding in Table 9.2.1. according to the age of the respondent either in 1999 or, if a coach, in the year the individual first registered as a coach. This coding mechanism reflects the age category of the individual at the latest time at interview of a decision to enter the Coach Education programme.

A related explanatory variable is $X_{29}$, “Gener”, the generation in which the individual is born. The difference between the variable “Gener” and the previous variable “Age” is that “Age” in this research is related to the date at which the latest decision is made to enter or not enter the Coach Education programme. Unlike “Age”, “Gener” is not effected by whether or not the individual is a coach. In line with “Age”, “Gener” is treated as a categorical explanatory variable as shown in Table 9.2.2.

<table>
<thead>
<tr>
<th>Category Code</th>
<th>Generation, Year of Birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>After 1970</td>
</tr>
<tr>
<td>2</td>
<td>1961-1970</td>
</tr>
<tr>
<td>3</td>
<td>1951-1960</td>
</tr>
<tr>
<td>4</td>
<td>1941-1950</td>
</tr>
<tr>
<td>5</td>
<td>1931-1940</td>
</tr>
<tr>
<td>6</td>
<td>1921-1930</td>
</tr>
</tbody>
</table>

Variable $X_{29}$, “Gener”, was allocated the Category code in Table 9.2.2. according to the year of birth of the respondent.

As discussed in Section 9.2.10., “Gener” may be correlated with the variable “Pay”, the attitude to paying coaches. It may also be correlated to the variable “Vol”, the attitude to others volunteering, discussed in Section 9.2.9.

The hypothesis related to potential explanatory variable $X_{28}$, “Age”, was that the probability of the individual deciding to volunteer as an athletics coach increases with each category from Age category 1 to Age category 3 and then decreases with each
subsequent category. The null hypothesis is that the probability of becoming a coach $(Y^* = 1)$ is not significantly associated with Age.

The hypothesis related to potential explanatory variable $X_{29}$, “Gener”, was that the probability of the individual deciding to volunteer as an athletics coach increases with each category from Generation category 1 to category 6. The null hypothesis is that the probability of becoming a coach $(Y^* = 1)$ is not significantly associated with Generation.

9.2.12. CLUB ASSOCIATION

Research related to this Section has shown that the probability of any individual choosing to volunteer may be associated with the ‘culture’ of the organisation to which the respondent is associated. The three National Surveys of Volunteering in the UK of 1981, 1991 and 1997 have shown that ‘being asked’ is consistently cited as being one of the key reasons for volunteering. The organisational ‘culture’ may effect the degree to which ‘asking’ takes place. The organisational ‘culture’ is likely to be related to the socio-economic and demographic composition of the club, as well as it’s history. In turn, these will be determined by the geographic area and the costs involved in membership.

In this research, two similar clubs have been sampled, labelled ‘A’ and ‘B’. Explanatory variable $X_{30}$ was set at “1” if the respondent is member of club ‘A’ or has a child or children who are members of club ‘A’ or, in the case of a coach, who first entered the Coach Education programme as a member of club ‘A’. Otherwise, $X_{31}$ was set at “0”.

The hypothesis is that probability of the individual deciding to volunteer as a coach increases if the individual is or was associated with club ‘A’ in contrast to club ‘B’. The null hypothesis is that the probability of becoming a coach $(Y^* = 1)$ is not
significantly associated with whether the individual is or was associated with club ‘A’ in contrast to club ‘B’.

9.3. Qualitative Confirmatory Information

An objective of the mathematical model developed in this research is to be able to categorise individuals into those who are likely to become coaches and those who are unlikely, given a number of quantitative facts about each individual. There are three main options to seek confirmation of the model:

i. To take another sample, use the model to estimate the probability of each respondent being a coach and to compare these estimates with the observed status of each respondent.

ii. To assess the state of each of the respondents used in the model after a period, say in the year 2010, apply the questionnaire again, use data in the model to re-estimate the probability of being a coach and compare the results with the observed status of each respondent.

iii. Where respondent’s circumstances are known to have changed and where respondents state they have ceased coaching or may coach in the future, to estimate new probabilities from model using changed circumstances and compare these probabilities with the statements of the respondent.

Option(i) was not available to the research due to insufficient time and resources to duplicate the interviewing. Option (ii) was not available to the research as it would require a time beyond the limits of the research. Also, given the turn over of members in athletic clubs, the success rate of follow up is likely to be low. For these reasons, Option (iii) was chosen. This was one of the main reasons for including qualitative questions in the interview.
Every non coach (parent and member) was asked “Might you ever take up athletics coaching yourself?” Where necessary, prompts were given such as “under any circumstances”, “if you present circumstances changed”, and “have you ever given it some thought?”. No time limit in the future was set. The question was asked almost at the end of the interview when the respondent was relaxed about the interview and with the interviewer. Prior to this question, the subject of coaching and the views of the respondent about it were raised in a series of open ended questions. The respondent had only been told that the research was about people and sport. They were deliberately not told that it was focused on coaching and entry to the Coach Education programme. In this way, it was hoped that respondents would be more likely to give an unbiased answer to the question as to whether they might enter the programme sometime in the future.

Respondents giving a positive reply were asked to state what changes of circumstance might lead them to coaching. The model could be re-run to give a new estimate for the probability of being a coach for each of these respondents altering the Explanatory variables corresponding to their statement of necessary changes in circumstances. A revised estimate would be obtained for the probability of being a coach under the changed circumstances stated by the respondent. If the model is satisfactory, it would be expected that a significant number of these revised estimates would be higher than the original estimates, confirming the positive replies of the respondents.

Respondents giving a negative reply were asked why they would not enter coaching. It was recognised that many would either use a phrase such as “under no circumstances” or decline to give any reasons or give replies that they felt were socially acceptable rather than the real reasons. The model could be re-run altering Explanatory variables corresponding to their statement of what circumstances deterred each respondent. It was expected that this would only be possible for a small number of respondents. This would have the effect of lessening the effects of those variables stated to be deterrents. If the model is satisfactory, it would be expected
that a significant number of these revised estimates would be higher than the original estimates, confirming the negative replies of the respondents.

The model can not be confirmed in a similar manner for those respondents who have taken and implemented the decision to enter the Coach Education Programme. The confirmation process described in the previous three paragraphs is based upon changing circumstances of individuals to increase their probabilities of becoming a coach. As existing coaches have already made that decision, the only decision then open to them is to cease coaching. Whilst some, if not all, of the variables in the model may be involved in this cessation decision, other explanatory variables are likely to be involved resultant from the experience of being a coach. An example is the perceived value of the relationships formed with the athletes they coach. Another example is the novelty value of being a coach. In addition, the process of being a coach may, in itself, alter the values of some of the Explanatory variables involved in the coach decision process. Examples are attitudes towards volunteering and towards payment of coaches.

The decision process to cease being a coach is a time extension of the Sports model for Coaching presented in Diagram 5.3.1. The vertical lines at decision points labelled “Join” and “Coach” can be replaced by “Enter Coaching and “Cease Coaching” respectively. In other words, the model is extended in time to take in a third decision, cease coaching. The decision to cease coaching fits in the concept of the Mintern model of decisions presented Diagram 5.2.2. The vertical lines of action points labelled “Resources Open” and “Drop Out” can be replaced by “Enter Coaching” and “Cease Coaching” respectively. Both the Sports Model and the Mintern model encompass the concepts of ‘key events’ and the Family Life Cycle propounded by Brandenburg (Brandenberg, J, and others, 1982) and Rapaport (Rapoport, R, and Rapoport, R N, 1975) discussed in Chapter 5.

As a consequence, if a logistic regression model can provide a satisfactory representation of the decision to become a coach, a similar model could be developed.
for the decision to cease being a coach. However, it was beyond the scope of this research to develop the model for cessation of coaching. A larger sample would be needed to gain sufficient information about individuals who had ceased coaching. More variables would have to be collected, requiring a further increase in sample size.

As ceasing coaching may be related to cessation of membership of an athletics club and also might be associated with negative personal events (illness, death, unemployment, family break down, etc.), obtaining a random sample might be difficult.

As a consequence, a different confirmation approach was taken. Each respondent defined as a coach was asked to say:

i. if they were, currently, a registered coach. Each statement was compared to the official Register of Coaches

ii. why they entered the Coach Education Programme

iii. in the light of their subsequent experience, why they might repeat the decision or might have taken the decision not to coach

iv. if they did not or had not coached at their club, whether they had coached elsewhere and the reasons

v. if they had ceased coaching, why they ceased and when

Confirmation of the model would be if the reasons given from (i) and (ii) could be associated with the explanatory variables in the model or, if not, the respondent as a coach did not fit the definition of "voluntary" as stated in Section 2.2. and the definition of "Registered Coach" as stated in Section 3.2. Situations not fitting these definitions would include:

- people who entered the programme primarily to obtain a qualification to assist in obtaining employment or entry to higher education
- professionals (teachers, leisure workers, physical education staff) paid to coach
- people who stated that they were coaches, but who had either not qualified under the Coach Education Programme as at 1999 or who had qualified but had subsequently had their qualification revoked or not renewed.
9.4. Background Information

A number of questions were included in the interview that were not intended to be used directly in the research. These questions were used for 3 reasons:

i. To construct a profile of the sample

ii. By using simple questions, to relax respondents

iii. To feedback to the participating clubs and to the County and National organisations in athletics information on other aspects of coaching which were known to be of interest:

- How and why a respondent became a coach
- Views on and perceptions of coaches and coaching
- Suggestions for improvements in coaching
- Present problems in coaching
- Coach education and revalidation

The profile of the sample is shown in Table 10.2.1. The feedbacks to clubs and to the associations are not included in this paper. All feedbacks will strictly respect the pledge of confidentiality given to every respondent and will be constructed in such a way that individual respondents can not be identified. The identity of the two clubs will not be revealed unless the Committees of both clubs agree in writing having seen drafts of any relevant documents.
10. THE DATA & MODELLING

10.1. Missing Data

As reported in Section 8.5., 112 respondents were interviewed. Complete sets of data for all potential explanatory variables were collected from 110 of these respondents. Data for 4 potential explanatory variables were not obtained from the remaining 2 respondents. This was because one respondent did not read any newspapers, books, magazines or journals and another respondent only read the headlines in newspapers occasionally. Both respondents were able to read and respond in writing to the printed Showcards used in the interviews. As a consequence and in both cases, no reliable estimates could be made for 4 potential explanatory variables associated with literacy and discussed in Section 9.2.8.:

i. \(X_{22}\), “Flesch”;
ii. \(X_{23}\), “Word”;
iii. \(X_{24}\), “Sent”;
iv. \(X_{25}\), “Science”

The result of this missing data was that two tables of data weightings had to be constructed. These are shown in Tables 8.5.1. and 8.5.2. The weightings in the first Table were used whenever data from all 112 cases was used. The weightings in the second Table were used whenever data sets were used which included in one of the variables listed in (i) to (iv) above.

10.2. Data Summaries

Table 10.2.1. summarises the quantitative information from all 112 respondents for all potential Explanatory variables and some additional descriptive variables. These are given by gender and, within gender, by coach and not coach (members and parents). Values are averaged over the sample, not the population. In interpreting these data, it
must be born in mind that the proportions of the sample to the estimated population differ between coaches, members and parents; between gender and between clubs. The actual proportions are given in Tables 8.5.1. and 8.5.2.

Table 10.2.1.
SUMMARY OF DESCRIPTIVE & POTENTIAL EXPLANATORY VARIABLES

<table>
<thead>
<tr>
<th></th>
<th>MALE</th>
<th></th>
<th>FEMALE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coach</td>
<td>Not</td>
<td>Coach</td>
<td>Not</td>
</tr>
<tr>
<td>Sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number interviewed</td>
<td>23</td>
<td>41</td>
<td>11</td>
<td>37</td>
</tr>
<tr>
<td>Club association</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count Club A</td>
<td>9</td>
<td>19</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>% Club A</td>
<td>39%</td>
<td>46%</td>
<td>64%</td>
<td>49%</td>
</tr>
<tr>
<td>Year of birth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average year</td>
<td>1946</td>
<td>1954</td>
<td>1949</td>
<td>1956</td>
</tr>
<tr>
<td>Year of Coach Registration</td>
<td>1986</td>
<td></td>
<td>1991</td>
<td></td>
</tr>
<tr>
<td>Age at Coach Registration</td>
<td>Average Age</td>
<td>40</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>Boys NOT in athletics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number in household</td>
<td>0.26</td>
<td>0.22</td>
<td>0.09</td>
<td>0.43</td>
</tr>
<tr>
<td>Boys IN athletics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number in household</td>
<td>0.39</td>
<td>0.20</td>
<td>0.36</td>
<td>0.22</td>
</tr>
<tr>
<td>Girls NOT in athletics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number in household</td>
<td>0.26</td>
<td>0.44</td>
<td>0.36</td>
<td>0.30</td>
</tr>
<tr>
<td>Girls IN athletics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number in household</td>
<td>0.70</td>
<td>0.27</td>
<td>0.73</td>
<td>0.43</td>
</tr>
<tr>
<td>Men NOT in athletics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number in household</td>
<td>0.09</td>
<td>0.15</td>
<td>0.73</td>
<td>0.62</td>
</tr>
<tr>
<td>Men IN athletics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number in household</td>
<td>0.04</td>
<td>0.02</td>
<td>0.27</td>
<td>0.22</td>
</tr>
<tr>
<td>Women NOT in athletics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number in household</td>
<td>0.61</td>
<td>0.73</td>
<td>0.09</td>
<td>0.19</td>
</tr>
<tr>
<td>Women IN athletics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number in household</td>
<td>0.30</td>
<td>0.12</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Distance to track</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average units</td>
<td>42</td>
<td>79</td>
<td>83</td>
<td>66</td>
</tr>
<tr>
<td>Minutes to track</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average minutes</td>
<td>13</td>
<td>21</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Cars available</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number</td>
<td>1.4</td>
<td>1.7</td>
<td>1.3</td>
<td>1.7</td>
</tr>
<tr>
<td>School leaving age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average year</td>
<td>16.7</td>
<td>17.0</td>
<td>17.0</td>
<td>17.1</td>
</tr>
<tr>
<td>Post school education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number of years</td>
<td>2.3</td>
<td>2.4</td>
<td>1.5</td>
<td>2.4</td>
</tr>
<tr>
<td>Hour home</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average hour, p.m.</td>
<td>5.81</td>
<td>5.21</td>
<td>5.23</td>
<td>4.64</td>
</tr>
<tr>
<td>Irregular working hours</td>
<td>Number with irregular work hours</td>
<td>10</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>% irregular working hours</td>
<td>43%</td>
<td>56%</td>
<td>9%</td>
<td>32%</td>
</tr>
<tr>
<td>Other sport, respondent</td>
<td>Number of respondents</td>
<td>14</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>% respondents with other sports</td>
<td>61%</td>
<td>54%</td>
<td>73%</td>
<td>70%</td>
</tr>
<tr>
<td>Other sport, dependants</td>
<td>Number of households</td>
<td>11</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>% households with other sports</td>
<td>48%</td>
<td>32%</td>
<td>45%</td>
<td>51%</td>
</tr>
<tr>
<td>Post school competition</td>
<td>Average % of years</td>
<td>66%</td>
<td>50%</td>
<td>54%</td>
</tr>
<tr>
<td>Interest in veteran athletics</td>
<td>Number interested</td>
<td>12</td>
<td>27</td>
<td>6</td>
</tr>
<tr>
<td>% interested in veterans</td>
<td>52%</td>
<td>66%</td>
<td>55%</td>
<td>43%</td>
</tr>
<tr>
<td>Main interest Road Running</td>
<td>Number main interest</td>
<td>6</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>% main interest in Road</td>
<td>26%</td>
<td>44%</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td>Participation in sport</td>
<td>1=at least 1/week, 5=None</td>
<td>1.5</td>
<td>2.0</td>
<td>1.8</td>
</tr>
<tr>
<td>Sport in work</td>
<td>Average % of work time</td>
<td>16%</td>
<td>12%</td>
<td>3%</td>
</tr>
<tr>
<td>Flesch Reading Score</td>
<td>Higher = Easier</td>
<td>51.5</td>
<td>51.5</td>
<td>54.7</td>
</tr>
<tr>
<td>Word complexity</td>
<td>Average syllables/word</td>
<td>1.55</td>
<td>1.57</td>
<td>1.51</td>
</tr>
<tr>
<td>Sentence length</td>
<td>Average words/sentence</td>
<td>23.0</td>
<td>22.6</td>
<td>20.8</td>
</tr>
<tr>
<td>Read science &amp; technology</td>
<td>Number who read science &amp; techn.</td>
<td>15</td>
<td>32</td>
<td>4</td>
</tr>
<tr>
<td>% read science &amp; technology</td>
<td>65%</td>
<td>78%</td>
<td>36%</td>
<td>35%</td>
</tr>
<tr>
<td>Attitude to volunteering</td>
<td>1=Str. Disagree, 5=Str. Agree</td>
<td>3.7</td>
<td>3.8</td>
<td>3.7</td>
</tr>
<tr>
<td>Attitude to paying coaches</td>
<td>1=Str. Agree, 5=Str. Disagree</td>
<td>3.6</td>
<td>2.7</td>
<td>3.8</td>
</tr>
</tbody>
</table>

NOTES

(1) Data for men and women in household refer to numbers EXCLUSIVE of the respondent.

(2) Time to club refers to respondent's perception and response, not to actual time.

DATA & MODELLING
(3) Percentages are of sample, not of population.

The following tendencies may be seen in the data summary in Table 10.2.1. No attempt is made here to statistically justify or verify any of these apparent tendencies:

a) Male coaches tended to have first registered in earlier years than female coaches.
b) Coaches tend to have more children involved with athletics than non coaches.
c) All respondents tend to have more girls in their households than boys.
d) The perceptions of the time required to get to the track from home are not related to the actual distances involved.
e) Coaches tend to have access to fewer cars than non coaches.
f) Coaches tend to have left school earlier than non coaches.
g) Coaches tend to have less post 15 formal education than non coaches.
h) Coaches tend to be home later than non coaches.
i) Coaches tend to have more regular hours than non coaches.
j) Coaches tend to be involved less with sports other than athletics than non coaches.
k) Households of coaches tend to be involved more with sports other than athletics than households of non coaches.
l) Coaches tend to have spent a greater proportion of their working life participating in sport than non coaches.
m) Coaches tend to be less interested in road running than non coaches.
n) Even at registration, coaches tend to be participating more frequently in sport than non coaches.
o) Coaches tend to read easier material than non coaches.
p) Coaches tend to disagree more with paying coaches than non coaches.

This preliminary data screening indicates that there appear to be different data patterns for coaches to non coaches. This gives some reassurance that a statistical model might be feasible to distinguish between coaches and non coaches.
The next two Sections, 10.3. and 10.4., describe the processes that were used to ensure that the data conformed to two of the most important assumptions of logistic regression: that the relationships of the explanatory variables with the response variable are linear and that the explanatory variables are independent of each other.

10.3. Data Vetting

A basic assumption of all linear regression models, including logistic regression, is that the relationship is linear throughout each continuous explanatory variable in relationship to the response variable. That is that one unit change of the explanatory variable causes the same change in the response variable throughout the range of the explanatory variable. If a non-linear relationship exists between the collected data of an explanatory variable and the response variable, then the collected data must be transformed to regain linearity before that explanatory variable is used to create a logistic model.

Two methods of transformation are possible. One method is to apply a mathematical transformation to all the data. Typically the transformation is the square root, inverse or logarithm if the relationship is a simple curve. The problem with using such mathematical transformations is one of interpreting the results and justifying the relationship in the real world.

The second method is to transform the variable from a Continuous form to a Categorical form. In essence, this transformation represents the non-linear continuous data by a series of straight lines, one within each category, with a step link between each category. This step relationship can be handled by logistic regression as a categorical variable. The variable is treated as a linked series of binary (2 state) variables. Each category in the series is compared to one category, the reference category. The interpretation of the logistic regression coefficients is that each coefficient measures the size and direction of the step between the category and the reference category.
Categorical transformation can be illustrated as follows. Supposing the number of cars in a household is found to be associated with the response, estimated probability of being a coach. This association is illustrated in Diagram 10.3.1.

Diagram 10.3.1.

ASSOCIATION OF NO. OF CARS WITH HOUSEHOLD SIZE (RESPONSE)

Whilst it may be possible to fit a curve to these points, the curve may have little meaning in reality and not be very instructive. In this case, fractions of cars have no meaning. The variable is discrete with 7 categories (0,1,2,3,4,5 and 6). However, some of the categories related to this variable would have few entries. For example,
there was only 1 case of 6 cars. Computational problems are likely from the under dispersion.

Looking at the diagram, the response appears to decline between 1 and 2 cars, declines even more between 2 and 3 cars and then remains steady for 3 or more cars. This interpretation can be illustrated by redrawing the association graph as shown in Diagram 10.3.2. In this, each group of similar observations (0 and 1 car, 2 cars, 3 and more cars) is represented by a straight line against the response. Each of the 3 lines is labelled can be represented as a category in a Discrete Variable. Using 3 categories reduces the risk of computational problems. Categories of categorical variables are introduced into logistic regression by use of dummy variables.
The significance and use of dummy variables will be shown later in this Section.

Referring to Diagram 10.3.2., the condition of 0 or 1 car is represented by dummy variable 1 having the value of 1, otherwise dummy variable 1 has a value of 0. Similarly, the condition of having 2 cars is represented by dummy variable 2 having a value of 1 (otherwise 0) and the condition of having 3 or more cars by dummy variable 3 having a value of 1 (otherwise 0). Taking all 3 dummy variables, the 3 conditions can then be represented as follows:

<table>
<thead>
<tr>
<th>Dummy Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 or 1 car</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 cars</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3 or more cars</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

This coding then allows the response to 2 cars (dummy variable set having a value of 010) to 0 or 1 cars (value 100) and the response to 3 or more cars (value 001).

In the case of this research, categorical transformations are indicated by many explanatory variables as the data is, in fact, measured in discrete steps. Examples are numbers of cars; numbers of children or adults; and year left school.

In conventional linear regression, linearity may be detected by visual inspection of graphical plots of the data pairs of the observed values an explanatory variable and the observed values of the response variable. In logistic regression, the observed response variable is binary. A straight graphical plot of observed values can only indicate whether or not the set of values of the explanatory variable paired with the observed response value of ‘0’ appear to differ from the set of values paired with the observed response value of ‘1’. It gives no indication on the shape of the relationship.
However, the response variable in the logistic regression model is the logit of the response variable. If observed values of an explanatory variable are grouped together, then the logit of the observed values of the response variable for each group can be calculated. A graph can then be constructed of the average of the observed values in each group of the explanatory variable and the logit of the observed values of the response variable for each group. Visually, this graph may indicate departures from a linear relationship.

There are two problems in constructing such graphs. If the average of the response variable for a group is 0.000, that is all the observed values are ‘0’, or the average is 1.000, that is all the observed values are ‘1’, then a value for the logit can not be computed. In such cases, the values 0.0001 and 0.9999 are substituted respectively for 0.000 and 1.000.

Table 10.3.1. shows an example of the calculations for the explanatory variable, number of cars available (Cars). The results are shown graphically in Diagram 10.3.3.

The results for cars shown in the previous Table and Diagram indicate that there may be a relationship between the explanatory and response variables. But the indications are that the relationship may not be linear.

<table>
<thead>
<tr>
<th>Number of cars</th>
<th>Frequency</th>
<th>Group Mean</th>
<th>Adjusted Mean</th>
<th>I it</th>
<th>logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>0.5000</td>
<td>0.5000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>1</td>
<td>51</td>
<td>0.3922</td>
<td>0.3922</td>
<td>-0.43825</td>
<td>-0.43825</td>
</tr>
<tr>
<td>2</td>
<td>49</td>
<td>0.2653</td>
<td>0.2653</td>
<td>-1.01857</td>
<td>-1.01857</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0</td>
<td>0.0001</td>
<td>-9.21024</td>
<td>-9.21024</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0.0001</td>
<td>-9.21024</td>
<td>-9.21024</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0.0001</td>
<td>-9.21024</td>
<td>-9.21024</td>
</tr>
</tbody>
</table>
Initially, the variable was transformed into a categorical variable with three values:

- Carscat = 0 when Cars = 0 or 1
- Carscat = 1 when Cars = 2
- Carscat = 2 when Cars = 3 or more

It was also noted that the relationship may be negative. The more cars reported, the less likely that respondent is likely to be a coach. This is the opposite to the hypothesis put forward in Section 9.2.4.

One explanation may be that larger numbers of cars are associated with larger households. The Pearson Correlation Coefficient between the total number of people in each household and the number of cars available was 0.345 (n=112). This low coefficient, combined with the shape of the distribution of pairs in a scatter graph, indicates that any relationship between household size and number of cars is very weak. Alternatives to the size of household are the number of adults, the number of children and the child/adult ratio. All of these produce similar evidence of only a weak relationship.

Alternatively, the number of cars may be related to the economic wealth or socio-economic grouping of the household. A negative relationship with being an athletics coach would then indicate that there may be an association between being a coach and
income or socio-economic grouping. Such an outcome would be in contrast to most research on volunteering in general that has indicated the opposite relationship.

Number of cars might also be associated the variety of outside interests of the household. The more the variety, the more competition there may be for transport to access these outside interests. One question in the interview was whether or not dependants were involved with sports other than athletics. This is a binary variable. The technique can be used again of plotting the logit of this binary variable, ‘other’, against the explanatory variable, ‘cars’. Using a categorical form of ‘cars’, the results are shown in Diagram 10.3.4. This indicates that such a relationship may exist, the probability of having other sporting interests increasing with the number of cars in the household.

**Diagram 10.3.4.**

**CARS V. PROBABILITY OF OTHER SPORTS**

Diagrams 10.3.5. to 10.3.30. are the resultant graphs for each of the other continuous potential Explanatory variables. The associated Tables have been omitted as some of them are large. Any significant Table contents are mentioned in the commentary following each pair of Diagrams.

How far the respondent has to travel to reach the club track is now considered, in terms of time and of distance.
Comparison of Diagram 10.3.5. and Diagram 10.3.6. confirms that there is no simple relationship between actual distance and reported (perceived) time from home to track. Diagram 10.3.5. indicates that there may be a linear relationship between time and the response variable. Inspection of the data Table indicated that the relationship may be binary. A cut off point of 6p.m. was chosen as weekday coaching sessions at clubs start at 6.30 p.m. The variable was transformed as follows:

\[
\text{Timebin} = \begin{cases} 
0 & \text{when \ Time < 6.01} \\
1 & \text{when \ Time > 6.00} 
\end{cases}
\]

It was noted that any relationship may be negative. The longer a respondent reports the time to reach the track from their home, the less likely that respondent is to be a coach. This is in accordance with the Hypothesis put forward in Section 9.2.3.

The earliest time is now considered at which the respondent is home and free from work on a Tuesday or Thursday evening, normal coaching days. The results of the interviews tended to cluster around each hour with a minor cluster around each half hour. For these reasons, the data is represented as recorded at the interview in Diagram 10.3.7. and in coarse groups in Diagram 10.3.8.
Diagram 10.3.7. indicates that there may be a weak positive relationship between the time at which a respondent is free at home during the week and the probability of being a coach. This positive relationship is in the opposite direction to that hypothesised in Section 9.2.6. By transforming the variable, Free, to a categorical form, Freecat, this relationship is more clearly shown in Diagram 10.3.8. It would appear that the relationship is not only positive, but it is also non linear, the probability rising more sharply between 6 p.m. and 8 p.m. than between 4 p.m. and 6 p.m. For model building, the variable was transformed into a categorical with values as follows:

- Freecat = 0 when Free < 4.01 p.m.
- Freecat = 1 when Free = <6.01 and > 4.00 p.m.
- Freecat = 2 when Free = >6.00 p.m.

Possible explanations of any positive relationships found in the model are discussed in Section 10.9.

The relationships with the number of dependants are now considered, starting with boys under the age of 18, divided between those who are not members of an athletics club and those who are members, shown in Diagrams 10.3.9. and 10.3.10. respectively.
As anticipated in the hypotheses put forward in Section 9.2.2., there is a negative relationship between the probability of being a coach and the number of male dependants under 18 not in an athletics club and a positive relationship with the number in an athletics club. However, the relationships are not linear. In the case of boys not in an athletics club, the probability of being a coach declines more rapidly when there are two than when there is only 1. In the case of boys who are members of an athletics club, the probability of being a coach rises sharply from 0 to 1 boy, but the increase continues less sharply with 2 boys. Both shapes are intuitively acceptable. As the variables are, strictly, categorical, they are entered in this form in the model.

The same analysis is conducted for girls under the age of 18 in Diagrams 10.3.11. and 10.3.12.
Girls appear to have similar relationships to those of boys with the probability of being a coach. Comparison of Diagram 10.3.9. and 10.3.12. indicates that having girls involved with athletics may have a stronger relationship with being a coach than having boys involved. Note the higher and positive values of the logit for girls. Again, both variables are treated as categorical for entry into the model.

The analysis is repeated for men of 18 and over in the household in Diagrams 10.3.13. and 10.3.14.

Once again the relationships appear to be non-linear. But, in contrast to boys and girls, the directions appear to be reversed of the relationship with the probability of being a coach, rising in the case of men not involved and declining with men involved. However, the graphs mask the fact that there were only 5 cases for 2 and 3 men not involved and only 1 case for 3 men involved. If these extremes are ignored, the effects of both variables on the probability of being a coach appear to be very weak. Both variables, MNot and MYes respectively, were transformed into binary variables as follows:

\[
\text{MNotbin} = 0 \text{ when } \text{MNot} = 0 \quad \text{MNotbin} = 1 \text{ when } \text{MNot} > 0
\]

\[
\text{MYesbin} = 0 \text{ when } \text{MYes} = 0 \quad \text{MYesbin} = 1 \text{ when } \text{MYes} > 0
\]

The analysis is repeated for adult female dependants in Diagrams 10.3.15. and 10.3.16.
Again, the relationships do not appear to be linear. Again, this may be due to there being few cases at the extreme, 4 cases of 3 women not involved and 1 case of 2 women involved. If these extremes are ignored, there may be a very weak negative relationship for women not in athletics and a very weak positive relationship for women involved. Both variables, FNot and FYes respectively, were transformed into binary variables as follows:

\[
\begin{align*}
\text{FNotbin} &= 0 \quad \text{when FNot} = 0 \\
\text{FNotbin} &= 1 \quad \text{when FNot} > 0 \\
\text{FYesbin} &= 0 \quad \text{when FYes} = 0 \\
\text{FYesbin} &= 1 \quad \text{when FYes} > 0
\end{align*}
\]

Diagrams 10.3.17. continues the vetting for the variable concerned with the proportion of work years in which the respondent had competed in sport. As the interview data tended to polarise with 20% of respondents at or near 0.0 and 20% at or near 1.0, the data was categorised into groups of 0.20 proportion sizes. The revised results are shown in Diagram 10.3.18.
Using the categorical diagram, 10.3.18., it would appear that there is a positive relationship. This is in accordance with the hypothesis in Section 9.2.7. But the relationship does not appear to be linear. A relatively small proportion of competition is associated with an increase in the probability of being a coach. This probability remains constant until over 80% of years have involved competition in sport when the probability slightly increases. The variable, ‘Compete’, was transformed initially into a categorical variable as follows:

Compcat = 0 when Compete <0.21  
Compcat = 1 when Compete >0.20 and <0.81.  
Compcat = 2 when Compete >0.80

Diagram 10.3.19. illustrates the linearity check conducted on the data of the response of respondents to the question “During the last 12 months, how often were you helping with and or competing in any formal organised sport or exercise (other than recreational walking)?” Responses were on a 5 point Likert scale where 1 = At least once a week and 5 = Not at all.
In Diagram 10.3.19., the three data points marked with * were each from only one data point. If these three outlying points are ignored, there appears to be a very weak decline in probability as the scale increases. That is, as the rate of participation declines, the probability of being a coach also declines. This is consistent with the hypothesis advanced in Section 9.2.7. However, the decline may not be linear. In view of the polarisation and the fact that nearly 70% of respondents marked ‘1.0’ on the scale, the variable, ‘Part’, was transformed to a binary variable as follows:

\[
\text{Partbin} = 0 \quad \text{when } \text{Part} < 2.51 \\
\text{Partbin} = 1 \quad \text{when } \text{Part} > 2.50
\]

The check on variable relating to association with sport was continued with the variable recording the proportion of years competing in any sport since leaving school. This is illustrated in Diagrams 10.3.20. and 10.3.21., the latter being a categorical transformation of the data.
No pattern can be readily identified in Diagram 10.3.20. This is due to the fact that most of the data points arise from very few cases, often only one. Diagram 10.3.21. is derived from grouping the data in the original variable, ‘Work’ into a new variable, ‘Workrev’, with groups as follows:

- Workrev = 1 when Work $\leq 0.00$
- Workrev = 1 when Work $<0.50$ and $<0.00$.
- Workrev = 2 when Work $>0.49$

Diagram 10.3.21. indicates that there is a positive and linear association between the probability of being a coach and the amount of sport in work over the previous 12 months. This accords with the hypothesis in Section 9.2.7. The new variable, ‘Workrev’ was used for model building.

The vetting was then continued to those variables concerning education. Diagrams 10.3.22. and 10.3.23. illustrate the relationship with probability of being a coach with the age at which the respondent left school.
In Diagram 10.3.22, the data, ‘School’ is presented as reported by the respondent. No pattern is revealed as both the first point (-1) and the last point (+4) result from very few cases. The data is, strictly, categorical as the units of reporting were complete years only. In Diagram 10.3.23, the data is transformed by grouping into a new categorical variable, Schoolcat’ as follows:

- Schoolcat = 0 when School < 1 i.e. left school by 15
- Schoolcat = 1 when School = 1 i.e. left school at 16
- Schoolcat = 2 when School = 2 i.e. left school at 17
- Schoolcat = 3 when School > 2 i.e. left school after 17

Diagram 10.3.23. indicates a non linear, negative association between the age of leaving school and the probability of being a coach. The probability drops sharply from 0 (left by 15) to 1 (left at 16). The rate of decline the slows to 2 (at 17) and slows more to 3 (after 17). This negative trend is contrary to the hypothesis put forward in Section 9.2.5. However, the Scottish survey of coaches (Lyle, J; Allison, M, and Taylor, J, 1997), found that coaches in athletics had significantly fewer years of formal education than coaches in other sports. The result reported here in this research may accord with the Scottish findings. The categorical variable, ‘Schoolcat’, was used in the model building.
Similarly, the data for years of Further and Higher Education is examined in Diagrams 10.3.24. and 10.3.25.

Diagram 10.3.24. LINEARITY CHECK - YEARS FE/HE

Diagram 10.3.25. LINEARITY CHECK - YEARS FE/HE (Categorical)

Diagram 10.3.24. of the reported data, ‘Fehe’, indicates that there may be a negative association. But any relationship is obscured because the number of cases tend to polarise around 0 and the first degree duration of 3 or 4 years. Additionally, the data is partially categorical, in the case of full time courses occupying complete academic years, and partially continuous from part time courses. Diagram 10.3.25. shows the same data as a categorical variable, ‘Fehecat’, transformed as follows:

\[
\begin{align*}
\text{Fehecat} &= 0 & \text{when Fehe} & < 2.00 \\
\text{Fehecat} &= 1 & \text{when Fehe} & > 1.99, <4.00 \\
\text{Fehecat} &= 2 & \text{when Fehe} & > 3.99
\end{align*}
\]

A non linear, negative trend is shown in Diagram 10.3.25. The probability of being a coach declines slightly from less than 2 years FE/HE to up 4 years (i.e. a first degree). The probability declines sharply after this. This trend is contrary the hypothesis put forward in Section 9.2.5., but continues the trend found in Diagram 10.3.22. for age of leaving school.

The next check was conducted on those variables concerned with reading level. The variable ‘Flesch’ measured the Flesch Ease of Reading Score for the newspaper or
magazine most frequently read by each respondent, taking into account whether or not they read articles mainly concerned with facts and/or articles mainly concerned with people and what they said. Diagram 10.3.26. illustrates the results.

**Diagram 10.3.26.**

*LINEARITY CHECK - FLESCH EASE OF READING SCORE*

The data point marked with * represents data from a single respondent. If this outlier is ignored, there appears to be a positive trend, which may be linear. The indication is that the probability of being a coach increases as the Flesch Score increases. As high Flesch scores are associated with easier reading, the Diagram indicates that the probability of being a coach decreases as reading level, as measured by the Flesch Score, decreases. This is the opposite effect to that hypothesised in Section 9.2.8.

However, research has indicated that Flesch Scores and School Grade are related (Flesch, R F, 1948). It could follow that School Leaving Age and Flesch Reading Score after leaving school are related. Calculation of the correlation between the variable ‘Flesch’ and the variable ‘School’ gives a Pearson Correlation Coefficient of 0.803 (n=110) indicating that a significant and positive correlation does exist in the data. Diagram 10.3.23. showed a negative, possibly linear, relationship between the probability of being a coach and School Leaving Age.

The tests for Collinearity reported in Section 10.4. did not reveal critical values for either ‘School’ or ‘Flesch’. This would indicate that any relationship between these
two variables does not significantly effect their correlation coefficients in a model with observed probability of being a coach as the response variable. However, the value of the variance-decomposition proportion in the correlation diagnostics of SPSS gave a value of 0.94 at the last step for ‘Flesch’. This indicates that there may be no high values for the variance-decomposition proportion were calculated for ‘School’.

As the data does not indicate a non-linear relationship, the continuous variable is used in model building as recorded. Further checks were made for non-linearity after the model was built.

Flesch scores are computed from two other variables that were measured: Word Complexity (‘Word’) measured as syllables per word; and Sentence Length (‘Sent’) measured as words per sentence. The analyses for these two variables are shown in Diagrams 10.3.27. and 10.3.28.

Word complexity appears to indicate a slight positive trend and sentence length a slight negative trend. In neither case do the diagrams indicate that the relationship is non linear. Both variables were used in model building without transformation.

Finally, the two attitudinal variables were tested. Diagrams 10.3.29. and 10.3.30. illustrate the check for variable recording the response to the statement: “People should volunteer some of their time to people elsewhere”. The responses were on a 5 point Likert scale.
Diagram 10.3.29 indicates that there may be a negative association between the attitude to others volunteering and to the probability of coaching. This result accords with the hypothesis put forward in Section 9.2.9.

The trend and the nature of it are made clearer by transforming the variable, ‘Vol’, into a categorical variable, ‘Volcat’. Diagram 10.3.30. shows the results. There is a non-linear negative association, declining sharply from the ‘strongly agree’ response to the ‘agree’ response. The appears to cease after this. As a result, the variable is transformed into a binary state with values as follows:

The values of the categorical variable, Volcat, are as follows:

\[
\begin{align*}
\text{Volbin} = 0 & \quad \text{when } \text{Vol} < 2.51 \\
\text{Volbin} = 1 & \quad \text{when } \text{Vol} > 2.50 \\
\end{align*}
\]

The binary variable, Volbin, was used in model building.

The final check was made on the variable recording the response to the statement: “Coaches at local sports clubs, generally, should be paid”. The responses were on a 5 point Likert scale. The results are shown in Diagram 10.3.31.

DATA & MODELLING
Diagram 10.3.31. indicates a positive association between the degree of disagreement with the statement to the probability of being a coach. This is in accord with the hypothesis put forward in Section 9.2.10. As there is only weak evidence that the trend is not linear, the continuous variable, Pay', is used initially in the model building.

**10.4. Initial Collinearity Test**

The second main assumption made in logistic regression modelling is that all the explanatory variables used in the model are independent of each other. That is to say that the effect on the response variable of one explanatory variable is completely independent of the effect of any other explanatory variable.

This independence was tested for by calculating the collinearity existing between each of the potential explanatory variables and the response variable, “Coach” and by inspecting the correlation matrix and variance decomposition proportions produced during the calculation. As shown and discussed in Section 7.5., unacceptable collinearity is present if the Tolerance statistic for any explanatory variable against the response variable has a value of 0.20 or less. One assumption of logistic regression is that all explanatory variables are independent of each other (see Section 7.2.2.1.). This assumption can be reviewed by inspecting the correlation matrix for values and the variance decomposition proportions for values exceeding 0.50. The Tolerance
statistics, the correlation matrix and variance decomposition proportions were
calculated for all potential explanatory variables and the response variable “Coach”
using the statistical functions within the linear regression function under Analysis in
SPSS. A constant term was allowed in the calculation.

Five potential variables had Tolerance statistics below 0.20:

- Age at decision
- Generation at decision
- School leaving age
- FEHE years
- Education years from 15

A further 5 potential variables had low Tolerance statistics and high correlations:

i. Word complexity
ii. Sentence length
iii. Flesch Ease of Reading Score
iv. Time to track
v. Distance to track

These findings were to be expected. There are logical associations between the values
of “Age at decision” and of “Generation”. “Flesch score” is derived from “Sentence
length” and “Word complexity”. There are logical associations between the values of
“Minutes to track” and “Distance to track”.

As a result of this analysis:

i. “Age at decision” and “Generation” were omitted from all model building.
ii. When “Flesch score” was included in the model building, both “Sentence
length” and “Word complexity” were omitted. If “Sentence length” was
included in the model building, “Word complexity” was omitted and vice
versa.

DATA & MODELLING
iii. When "Minutes to track" was included in the model building, "Distance to track" was omitted and vice versa.

Calculations of Tolerance statistics were repeated using decisions (i) to (iii) above. The Tolerance statistics and covariance matrices confirmed that no serious correlations or covariances existed with data sets conforming to these decisions.

10.5. Model Building

Following the decisions in Section 10.4., 25 or 26 potential variables still remained. The ideal method to build a model would be to enter all potential variables and allow the computer program to automatically decide which variables were significant at a given level of probability. This is not feasible for two reasons.

i. Combinations of variables should not be used if they cause significant collinearity as this would break an assumption of logistic regression (see Section 7.5). No programme is known that automatically carries out tests for collinearity.

ii. The number of variables and the data density of each variable may produce matrices that are either too large or too sparse for the programme to proceed.

As a consequence, it was decided to build an initial model by introducing one variable at a time, starting from the variable that appears to be most significant. Within logistic regression in SPSS, a number of alternative methods are available to introduce explanatory variables into the model. One of these is stepwise Forward selection using likelihood ratio criterion (Forward:LR) (SPSS Inc, 1999). The method starts with only a constant (cut off) in the model and adds one variable at a time. The selection process for this variable calculates the log-likelihood of a model with and without each potential variable in it. The ratio of these two log-likelihoods is
calculated and tested against the null hypothesis that the ratio is not significantly different from 0. The test is made by comparing the chi-square value for the bigger model to the chi-square for the reduced model (without the variable under test). This is a test of goodness of fit, which was discussed in Section 7.3. The test value is used as a score. The explanatory variable with the highest score is then added to the model. The variables in the revised model are then examined in the same way to see if any of them can be removed. The process is repeated with the remaining potential variables step by step until no further variables can be entered or removed from the model.

The default minimum probability value of 0.05 was chosen to allow variables to enter the model and the default probability value of 0.10 was chosen to eliminate variables. As one of the reading variables was expected to be included, the data weighting used was as per Table 8.5.2. (see Section 10.1.) A cut off value of 0.5 was used, that is all cases predicted to have a response value less than 0.5 will be classified by the model as “Not Coach” and all cases with values of 0.5 or greater will be classified as “Coach”.

The SPSS defaults were used of including a variable in the model and using Indicator coding for categorical variables, with the last category being used as the reference category.

The data was weighted according to Table 8.5.2. for n = 110. This reduced set of cases was used as the set of potential explanatory variables used at this stage included those concerned with reading levels, values for which were not available for 2 of the full 112 cases in the study.

Step 14 from this building procedure was chosen as the Starting Model. Subsequent steps were rejected as they either gave no further improvements in the number of correctly predicted cases or they gave a perfect fit. Perfect fits were rejected from all analyses as they indicate mathematical failures in the model building algorithm. Steps
giving no further improvements were rejected from all analyses on the grounds of parsimony (see Section 6.3.).

The Starting Model did not include any of the explanatory variables connected with reading levels. As a consequence, the same set of explanatory variables as this Starting Model were entered again into model building, now using the weights associated with all 112 cases as in Table 8.5.1. This was Starting Model (Revision 1)

The Starting Model (Revision 1) has a log likelihood value (-2LL) of 98.207. It contains 14 Explanatory variables and a constant. The computed Chi-square value to the Initial Model (one having only a constant with no explanatory variables) is 333.937. This is significant at a probability level less 0.001 (23 degrees of freedom). The Nagelkerke Correlation Coefficient ($R^2$) is 0.807, indicating that about 80% of the variability of the data is explained by the model. 100% of the “Not Coach” cases are correctly classified by the model. 87.1% of the “Coach” cases are correctly classified. All Tolerance statistics are above 0.20. These statistics indicate a satisfactory fit of the data explaining a high proportion of the variance with variables not significantly correlated to each other.

The computed value for the constant term was -30.757 with a standard error calculated at 82.397. The latter indicates that the computed value of the constant term is probably not significantly different from 0. This would accord with logic in that if there were no explanatory variables, the probability of an individual not becoming a coach would be zero.

The Starting Model (Revision 1) was run again without a constant in the model. The independence of the explanatory variables in this amended model was tested again. Four variables have Tolerance statistics below 0.20 indicating significant correlation with one or more of the other explanatory variables:

% Post school years competing  (0.199)
School leaving age (0.188)
Attitude to others volunteering (0.100)
Attitude to paying coaches (0.098)

Inspection of the correlation matrix and the variance decomposition proportions showed that there was a strong association between the values for the two attitudinal variables.

As a result, the model was re-run using the original set of potential explanatory variables less these two attitudinal variables. The constant term was re-introduced and the weights for n = 110 used again. The resultant Starting Model (Revision 2) had 11 variables in it, including one concerning reading levels.

It was observed that the constant term had a coefficient of -21.172 with a standard error of 37.451. This again indicated that the value of the coefficient of the constant term was not significantly different from zero. A model without a constant accords with logic in that, in the absence of any influencing variables, an individual would have equal probabilities (p = 0.5) of becoming and of not becoming a coach. If p=0.5, then the logit of the response variable = 0, that is the constant term in the logistic regression model will be zero.

All the variables in this model were tested to measure their significance to the model, in terms of the value of -2LL contributed. This was done by running the model without and with each variable and measuring the decrease in -2LL resulting from this. The model was also tested for the addition of each of the explanatory variables not included, again measuring the improvement in -2LL and the significance of this improvement. SPSS allows these tests to be conducted as variables can be entered in blocks and the results calculated for each block. The model variables are placed in Block 1. A variable is then either removed from this block and placed in Block 2 or the additional variable placed in Block 2.
The tests showed that one variable could be removed and one variable added. “Main interest road running” only contributed a -2LL amount of 0.986 that was not significant (1 degree of freedom, \( p=0.321 \)). “Number of Women in the household IN athletics” added 9.312 of -2LL that was significant (2 degrees of freedom, \( p=0.002 \)).

The model was re-run without a constant, without the road running variable and with the variable concerning women in athletics. This model was termed Starting Model (Revision 3). This model has a log likelihood value (-2LL) of 201.708. It contains 11 explanatory variables and no constant. The Nagelkerke Correlation Coefficient (R²) is 0.903, indicating that about 90% of the variability of the data is explained by the model. 100% of the “Not Coach” cases are correctly classified by the model. 60.8% of the “Coach” cases are correctly classified. All Tolerance statistics are above 0.20 (Menard, S, 1995; Hutcheson, G, and Sofroniou, N, 1999). Compared to Starting Model (Revision 1), the statistics indicate an improvement in fit with a reduced number of Explanatory variables.

Tests were again conducted to measure the effects of adding further variables. The variable concerning “Attitude to paying coaches” added significantly to the model, but caused some of the Tolerance values to fall below 0.20 indicating collinearity with other variables. The source of this collinearity was found by inspecting the correlation matrices and the variance decomposition proportions. These showed that the collinearity resulted from associations with the attitudinal variable on pay with variables associated with education, reading and amount of post school competition. Not allowing these variables to enter the model allowed the attitudinal variable on pay to be introduced with an improvement in the model using fewer variables. On the principle of parsimony, this reduced variable model was used. As it does not contain any reading variables, the data from all 112 cases could be used with the appropriate weighting. Again tests were conducted on the significance of the removal and addition of variables. It was found that the addition of the variable “Club” significantly improved the model and still kept the Tolerance statistics above 0.200. This was Starting Model (Revision 4).
Starting Model (Revision) 4 contains 10 categorical variables and 1 continuous variable, “Attitude to paying coaches”. This variable had been retained in its original continuous form as the graphical inspection in Diagram 10.3.30. did not show any obvious signs of non-linearity.

A quantitative method for testing for non-linearity is the Box-Tidwell transformation. This requires addition to the model of a new explanatory variable for the continuous explanatory variable tested. This new variable is formed as follows:

\[ X_{bt} = X^j \times \ln(X^j) \]  

(Equation 10.5.1.)

where:

- \( X_{bt} \) is the Box-Tidwell transformation of values of explanatory variable \( j \)
- \( X^j \) are the original values of explanatory variable \( j \)
- \( \ln(X^j) \) are the natural logarithms of the original values of explanatory variable \( j \)

(Hosmer, D W, and Lemeshow, S, 1989; Menard, S, 1995)

The Box Tidwell variable for “Attitude to paying coaches” was calculated and added to Starting Model (Revision 4). The variable added -2LL of 20.599 that is significant (1 degree of freedom, \( p<0.001 \)). It was concluded that the association of this variable with the response variable was not linear. As the variable resulted from responses to a 5 point Likert scale, the original data (Pay) was transformed into a categorical form (Paycat) as follows:

- Paycat = 0 when Pay <2.50, i.e. Agree with paying
- Paycat = 1 when Pay >2.49 and < 3.50, i.e. Neutral to paying
- Paycat = 2 when Pay >3.49 and < 4.50, i.e. Agree with paying
Paycat = 3 when Pay > 4.49, i.e. Strongly Disagree with paying

Tests on removing variables in the model and adding other variables not in the model indicated that either they did not contribute significantly or, if they did, they created significant collinearity, as measured by the Tolerance statistic.

Interactions with the explanatory variable "Gender" and other explanatory variables were then investigated. The correlation matrix of all the explanatory variables in Starting Model (Revision 4) was calculated. The Pearson Correlation Coefficient was used throughout as the objective was to identify major correlations, not to exactly measure them for significance. The results are shown in Table 10.5.1. Two variables with coefficients exceeding 0.20 are shown emboldened. The cut-off of 0.20 was taken as $r$ values above this are statistically significant ($\alpha = 0.05$) with a sample size of 112. (Black, T R, 1999)

<table>
<thead>
<tr>
<th>Explanatory variable in model</th>
<th>Correlation Coefficient to Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Club</td>
<td>0.046</td>
</tr>
<tr>
<td>Time to track</td>
<td>-0.117</td>
</tr>
<tr>
<td>Number of cars</td>
<td>0.068</td>
</tr>
<tr>
<td>Regularity of work hours</td>
<td>-0.227</td>
</tr>
<tr>
<td>Time home</td>
<td>-0.272</td>
</tr>
<tr>
<td>Number of Boys NOT in athletics</td>
<td>0.036</td>
</tr>
<tr>
<td>Number of Girls NOT in athletics</td>
<td>0.152</td>
</tr>
<tr>
<td>Number of Girls IN athletics</td>
<td>-0.079</td>
</tr>
<tr>
<td>Number of Women IN athletics</td>
<td>-0.171</td>
</tr>
<tr>
<td>Participation frequency</td>
<td>-0.053</td>
</tr>
<tr>
<td>Attitude to paying coaches</td>
<td>-0.079</td>
</tr>
</tbody>
</table>

Because of the significant correlations shown in Table 10.5.1., the introduction of both 'gender' and 'regularity of work hours' or of 'gender' and 'time home' would invalidate the basic assumption of logistic regression that all explanatory variables should be independent.

DATA & MODELLING
To cater for these correlations, interaction variables were formed with variable “Gender” (Gen) for the variables “Regularity of work hours” (Reg) and “Time home” (Freecat) as follows:

- **Genreg** = 0 when Gen = 0 and Reg = 0 Women with regular hours
- **Genreg** = 1 when Gen = 0 and Reg = 1 Women with irregular hours
- **Genreg** = 2 when Gen = 1 and Reg = 0 Men with regular hours
- **Genreg** = 3 when Gen = 1 and Reg = 1 Men with irregular hours
- **Genfree** = 0 when Gen = 0 and Freecat = 0 Women home by 4 p.m.
- **Genfree** = 1 when Gen = 0 and Freecat = 1 Women home after 4 p.m., by 6 p.m.
- **Genfree** = 2 when Gen = 0 and Freecat = 2 Women home after 6 p.m.
- **Genfree** = 3 when Gen = 1 and Freecat = 0 Men home by 4 p.m.
- **Genfree** = 4 when Gen = 1 and Freecat = 1 Men home after 4 p.m., by 6 p.m.
- **Genfree** = 5 when Gen = 1 and Freecat = 2 Men home after 6 p.m.

Each of these two interaction variables, “Genreg” and “Genfree” were introduced to Starting Model (Revision 4). Only “Genreg” added significantly to the model. This model was designated Starting Model (Revision 5).

The model was then tested again for the significance of variables in it and for the significance of adding other variables. All the variables in the model were significant. The added -2LL was significant when 4 other variables were introduced. 3 of these variables produced Tolerance statistics below 0.20 indicating that there was high collinearity. The explanatory variable “Number of Women NOT in athletics” contributed significant -2LL without creating low Tolerance statistics. This was added to the model to create Starting Model (Revision 6).
Tests were again conducted on the significance of the -2LL created by the deletion and addition of variables to Starting Model (Revision 6). 4 variables could be added with significant -2LL. But, in all cases, they created Tolerance statistics less than 0.20, indicating collinearity. It was concluded that no further improvements could be made to the model without violating basic assumptions regarding logistic regression models. Starting Model (Revision 6) was accepted as the Thesis Model.

The Thesis Model is based upon the data from all 112 cases. The details are shown in Table 10.5.2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>-2LL lost</th>
<th>R²</th>
<th>% correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thesis Model values</td>
<td>142.232</td>
<td>0.952</td>
<td>78.8%</td>
</tr>
</tbody>
</table>

Table 10.5.2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variable removed</th>
<th>-2LL lost</th>
<th>St. Error</th>
<th>R²</th>
<th>Degree Of Freedom</th>
<th>Sign</th>
<th>Toler.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Club (B)</td>
<td></td>
<td>5.077</td>
<td>1.149</td>
<td>26.987</td>
<td>1</td>
<td>&lt;0.001</td>
<td>0.351</td>
</tr>
<tr>
<td>Minutes to track (B)</td>
<td></td>
<td>11.592</td>
<td>11.126</td>
<td>9.313</td>
<td>1</td>
<td>0.002</td>
<td>0.590</td>
</tr>
<tr>
<td>Cars available (C)</td>
<td></td>
<td>10.510</td>
<td>2.310</td>
<td>86.052</td>
<td>2</td>
<td>&lt;0.001</td>
<td>0.391</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>4.503</td>
<td>1.743</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C)</td>
<td></td>
<td>13.062</td>
<td>3.367</td>
<td>36.797</td>
<td>3</td>
<td>&lt;0.001</td>
<td>0.403</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>9.299</td>
<td>2.500</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>13.760</td>
<td>3.302</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time home (C)</td>
<td></td>
<td>-8.735</td>
<td>2.021</td>
<td>36.841</td>
<td>2</td>
<td>&lt;0.001</td>
<td>0.336</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>-6.979</td>
<td>1.706</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys IN athletics (C)</td>
<td></td>
<td>-3.507</td>
<td>1.230</td>
<td>50.960</td>
<td>2</td>
<td>&lt;0.001</td>
<td>0.749</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>-3.221</td>
<td>1.369</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls NOT in athletics (C)</td>
<td></td>
<td>-12.481</td>
<td>11.096</td>
<td>22.908</td>
<td>2</td>
<td>&lt;0.001</td>
<td>0.598</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>-17.970</td>
<td>11.549</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls IN athletics (C)</td>
<td></td>
<td>-15.007</td>
<td>2.757</td>
<td>136.447</td>
<td>2</td>
<td>&lt;0.001</td>
<td>0.412</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>-2.871</td>
<td>1.446</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women NOT in athletics (C)</td>
<td></td>
<td>2.652</td>
<td>1.585</td>
<td>14.944</td>
<td>2</td>
<td>0.001</td>
<td>0.431</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>0.912</td>
<td>1.574</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women IN athletics (B)</td>
<td></td>
<td>-2.880</td>
<td>0.953</td>
<td>8.358</td>
<td>1</td>
<td>0.004</td>
<td>0.793</td>
</tr>
<tr>
<td>Participation (B)</td>
<td></td>
<td>4.084</td>
<td>0.968</td>
<td>25.084</td>
<td>1</td>
<td>&lt;0.001</td>
<td>0.500</td>
</tr>
<tr>
<td>Attitude; pay coaches (C)</td>
<td></td>
<td>-13.761</td>
<td>2.899</td>
<td>87.685</td>
<td>3</td>
<td>&lt;0.001</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>-8.686</td>
<td>2.330</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>-3.798</td>
<td>1.848</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONSTANT</td>
<td>NONE</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The Table shows that the Thesis Model has an initial log likelihood of 1464.176. The log likelihood of the model is 142.232 with 22 degrees of freedom giving a confidence value of <0.001 that the model significantly fits the data better than the initial model.
The Nagelkerke correlation coefficient $R^2$ is 0.952, indicating that about 91% of the original variation has been accounted for. The model correctly classifies all 1001 “not coach” cases. It correctly classifies 43 out of 55 “coach” cases, 78.8%. The calculated coefficients and their standard errors are shown in Table 10.5.2., together with the amount of log likelihood (-2LL) lost if the variable was removed, the degrees of freedom contributed and the significance. The final column shows the Tolerance statistic. NOTES to TABLE 10.5.2.

(B) = Binary variable
(C) = Categorical variable

(1), (2), (3) refer to the categories of the variable

The coefficient of a variable is with respect to the reference category of that variable. The reference category is the last category. No coefficient is given for the reference category as this is allocated values of ‘0’ in the coding.

A negative sign to a coefficient indicates that the variable decreases the probability of being a coach, in relation to the reference category.

The variables in Table 10.5.2. are now described

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUB</td>
<td>Club A in relationship to Club B</td>
</tr>
<tr>
<td>TIME</td>
<td>Home 20 minutes or less from home to track in relation to more than 20 minutes</td>
</tr>
<tr>
<td>CARS</td>
<td>(1) No or 1 car available in relation to 3 or more cars available (2) 2 cars available in relation to 3 or more cars available</td>
</tr>
<tr>
<td>REGULAR</td>
<td>(1) Male with regular hours of work in relation to a female with irregular hours of work</td>
</tr>
</tbody>
</table>

DATA & MODELLING
(2) Male with irregular hours of work in relation to a female with irregular hours of work

(3) Female with regular hours of work in relation to a female with irregular hours of work

HOME

(1) Home by 4 p.m. in relation to home after 6 p.m.

(2) Home after 4 p.m. and by 6 p.m. in relation to after 6 p.m.

BOYS IN

(1) No boys in household in athletic clubs in relation to 2 or more boys in athletic clubs

(2) 1 boy in household in athletic clubs in relation to 2 or more boys in athletic clubs

GIRLS NOT

(1) No girls in household in athletic clubs in relation to 2 or more girls not in athletic clubs

(2) 1 girl in household in athletic clubs in relation to 2 or more girls not in athletic clubs

GIRLS IN

(1) No girls in household not in athletic clubs in relation to 2 or more girls in athletic clubs

(2) 1 girl in household in athletic clubs in relation to 2 or more girls in athletic clubs

WOMEN NOT

(1) Other than respondent, no women in athletic clubs in relation to 2 or more women not in athletic clubs

(2) Other than respondent, 1 woman in household in athletic clubs in relation to 2 or more women not in athletic clubs

WOMEN IN

Other than respondent, no women in household not in athletic clubs in relation to 1 or more women in athletic clubs

PARTICIP.

Involved in sport at least once a month in relation to involved less than once a month

PAY

(1) Strongly agree with statement that “Coaches at local sports clubs generally should be paid” in relation to strongly disagree with statement.

(2) Agree with statement that “Coaches at local sports clubs generally should be paid” in relation to strongly disagree with statement.
Neutral to statement that "Coaches at local sports clubs generally should be paid" in relation to strongly disagree with statement.

Before accepting the Thesis Model, a number of diagnostic tests were made:

i. To measure the dispersion of the data used to build the model.

ii. To test that the fit of the model could not have come about by chance and that the classification by the model was significant.

iii. To investigate the influences on the model of each case and of each Explanatory variable.

These are reported in Section 10.6.

10.6. Confirmation & Diagnostics.

10.6.1. DISPERSION

Dispersion tests the assumption in logistic regression models that the variance of the response variable can be calculated from the mean of the distribution of observed values that variable as it is assumed to be a binomial distribution. Over or under dispersion occurs when the actual variance of the response variable is greater than or less than (respectively) the assumed value. The result of over or under dispersion is that the calculation of the coefficient values results in values that are too low (over dispersion) or too high (under dispersion). A test of dispersion is to calculate the statistic "s" as shown in Equation 7.4.2.

The model -2LL is 142.232 and the associated degrees of freedom are 22. From this, the value of s is 6.465. The fact that s is greater than 1.0 indicates that the data is over dispersed. The most probable reason for this over dispersion is that there are other significant explanatory variables that have not been included in the model.
The relative low Tolerance statistics for the variables in the model, shown in Table 10.5.2., indicate that these variables are indirectly and partially measuring other, probably complex, variables. This is partially confirmed by the fact that other variables do add significantly to the model, but also produce very low Tolerance statistics.

One example is the variable recording the age at which each respondent left school. Whilst this variable can add significantly to the model, it is correlated with several of the variables in the model, notably:

- Club \( (R^2 = 0.226) \)
- No. of women in athletics \( (R^2 = -0.339) \)
- Attitude to paying coaches \( (R^2 = 0.231) \)

Similar effects occur with the variable “Proportion of post school years competing in sport”; the variable recording attitude to others volunteering and the variables related to reading.

The effect of over dispersion is that the estimates of the coefficients are lower than the true values. However, the estimates of the coefficients are not biased with over dispersion. Their relative effects on the model remain the same. In Section 10.6. these relative effects are calculated and the results used to interpret the model.

The objective of the research was to prove that a logistic regression model was able to produce a significant fit to the variables suggested by past research. The over dispersion encountered is relatively low for a social based study and does not negate the objective of the research the accuracy of the calculated values for the coefficients, but does not alter their relative effects.
10.6.2. FIT & CLASSIFICATION ABILITY

The ability of the model to correctly classify is one measure of the fit of the model and indicates the possible power of the model to predict. However, it is perfectly possible for a random selection of variables to produce a good fit and classification. Two tests were applied to the Final Model: $\varphi_p$, the Proportional Change in Error; and the 'C' Statistic. These are discussed in Section 7.3. together with details of how they are calculated.

The Proportional Change in Error is based upon the classification of the cases. In the Final Model, 1004 cases who were not coaches were correctly assigned to “Not Coach” and no cases assigned to “Coach”. 47 cases who were coaches were correctly assigned to “Coach” and 4 cases incorrectly assigned to “Not Coach”. The weighting used was for an estimated population of 1,055. The resultant value of $\varphi_p$ was 0.955. The d test of this value has a value of 9.913 that exceeds the critical value of 2.581 ($\alpha = 0.01$). This indicates that the value of $\varphi_p$ is significantly greater than 0.0 and that the classification has a high reliability of being a result of the modelling and not of chance.

The ‘C’ Statistic also considers whether or not the classification could have occurred by chance. It examines the classification of each case and tests the probability that this classification could have been by chance (e.g. by tossing a coin). The statistic summarises all the pair tests. The statistic can range between 0.000 to 1.000. If the overall classification could have occurred by chance, C will have a value of 0.000. The higher the value of ‘C’, the less likely the overall classification could have occurred by chance. The ‘C’ Statistic for the Final Model is 0.952. This confirms that the classification is most unlikely to have occurred by chance.

Classification is determined by the chosen value of the cut off value. The Thesis Model uses the default value of 0.5. The effect of a change in the value of the cut off can be seen in Diagram 10.6.1. This plots the value of the response variable.
“Coach”, for each of the cases. The cases are ordered in the size of their predicted values. Those cases plotted to the right of the vertical 0.5 probability line are classified by the model as “Coach”. Those plotted to the left are classified as “Not Coach”. The symbols in the graph show whether of not each case is actually a coach or not.

A further indication of fit was made using the deviance values of the residuals (Equation 7.3.2.). Diagram 10.6.2. shows the distribution of the deviance values derived from the Thesis Model plotted with a straight line representing a normal distribution (SPSS Inc. 1999). The deviance distribution conforms to the normal distribution until the cases with higher probabilities associated with being a coach. The extreme end of the deviance distribution deviates substantially from the normal distribution.
There is a marked deviation from the normal where the predicted probabilities are moving towards 1.0, “Coach”. Possible reasons include:

i. One or more explanatory variables in the model that inadequately measure an effect.

ii. Exclusion of one or more significant explanatory variables.

The sample size and the correlations between variables did not allow further explanatory variables to be added, in particular interactions between variables.

As discussed in Section 7.3., the distribution of the deviance of binary data may not fully conform to a normal distribution. The plot shown in Diagram 10.6.2. is only indicative of the goodness of fit. Bearing in mind that the data is derived from a sociological study, the overall fit is considered to confirm the other statistics of fit and
classification which suggest the Thesis Model does produce a significant fit. Scott Menard (1995, p.79) in his work on quantitative applications in social science draws attention to this and says:

"Even cases with very large residuals ..... do not necessarily indicate problems in the model, insofar as we are dealing with nondeterministic models in which individual human choice and free will may naturally produce less than perfect prediction of human behaviour."

10.6.3. INFLUENTIAL CASES & VARIABLES

Whilst the data sample was taken using random methods of respondent selection, it is possible that there might be individual respondents with extreme characteristics as measured by the explanatory variables used in the Final Model. The data from any extreme cases may significantly influence the model. Any significant influence may effect the estimated value of the coefficient of one or more explanatory variables and/or the estimated value of the model constant.

One method of detecting cases with significant influences on the model is to calculate a value referred to as dBeta. This is derived from standardising a leverage function known as Cooke's Influence. The standardisation is obtained by dividing the value for Cooke’s Influence by the Leverage value. If dBeta is substantially greater than 1.0, then the deletion of this variable has a substantial influence and it’s deletion from the model building would substantially effect the estimates of the coefficients. Values of dBeta are normally distributed with a mean of 0.0 and a standard deviation of 1.0, (Menard, S, 1995).

10 cases have values of dBeta greater than 1.0 and are shown in Table 10.6.1., together with the probability calculated by them and the category from the model and observed actuality. Cases 110, 105, 59 and 103 are large. The other cases are not exceptionally high being under or just over 3.00 and might be expected in the normal
distribution from 112 cases. All 10 cases are actual coaches. The model classifies the first 8 as “not coach” and last two as “coach”.

Table 10.6.1. 

<table>
<thead>
<tr>
<th>Case No.</th>
<th>dBeta</th>
<th>Probability</th>
<th>Category Model</th>
<th>Category Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>468,116.4</td>
<td>0.000</td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>105</td>
<td>1,704.0</td>
<td>0.001</td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>59</td>
<td>787.8</td>
<td>0.001</td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>103</td>
<td>23.6</td>
<td>0.044</td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>54</td>
<td>3.2</td>
<td>0.310</td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>107</td>
<td>2.8</td>
<td>0.311</td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>62</td>
<td>2.2</td>
<td>0.373</td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>60</td>
<td>1.8</td>
<td>0.413</td>
<td>Not Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>112</td>
<td>1.7</td>
<td>0.506</td>
<td>Coach</td>
<td>Coach</td>
</tr>
<tr>
<td>64</td>
<td>1.2</td>
<td>0.571</td>
<td>Coach</td>
<td>Coach</td>
</tr>
</tbody>
</table>

These 10 cases correspond to the 10 cases at the right of the graph of Deviance values in Diagram 10.6.2. This indicates that these cases explain some of the lack of fit of the Final Model.

The influence by case of individual explanatory variables and of the constant can be measured by calculating the statistic dfBeta for each variable and case. dfBeta is the change in the estimated value of the coefficient for a variable when the case is excluded from the model. It indicates the effect on the estimated value of the coefficient of deleting that variable (SPSS Inc, 1999).

The dfBeta values for the 4 highest dBeta cases shown in Table 10.6.1. are shown in Table 10.6.2. Only in Case 110 do some values of dfBeta exceed the value of the estimated Standard Error of the Coefficient. These are indicated by the shading in the table.
Table 10.6.2.
VALUES OF $\text{dbeta}$ FOR CASES WITH LARGE VALUES FOR dBETA

<table>
<thead>
<tr>
<th>Variable</th>
<th>St. Error</th>
<th>$\text{dfBeta}$ for Case No.</th>
<th>110</th>
<th>105</th>
<th>59</th>
<th>113</th>
</tr>
</thead>
<tbody>
<tr>
<td>Club (B)</td>
<td>1.149</td>
<td>-1.746</td>
<td>-1.053</td>
<td>-0.109</td>
<td>-0.035</td>
<td></td>
</tr>
<tr>
<td>Minutes to track (B)</td>
<td>11.126</td>
<td>-5.022</td>
<td>-1.373</td>
<td>0.790</td>
<td>1.622</td>
<td></td>
</tr>
<tr>
<td>Cars available (C) (1)</td>
<td>2.310</td>
<td>-4.132</td>
<td>-1.578</td>
<td>0.256</td>
<td>0.689</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>1.743</td>
<td>-2.294</td>
<td>-1.088</td>
<td>0.888</td>
<td>0.676</td>
<td></td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C) (1)</td>
<td>3.363</td>
<td>-6.305</td>
<td>-2.834</td>
<td>1.210</td>
<td>1.491</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>2.500</td>
<td>-3.574</td>
<td>-1.656</td>
<td>-0.124</td>
<td>0.449</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>3.302</td>
<td>-6.290</td>
<td>-2.435</td>
<td>0.432</td>
<td>1.589</td>
<td></td>
</tr>
<tr>
<td>Time home (C) (1)</td>
<td>2.021</td>
<td>3.771</td>
<td>1.477</td>
<td>-0.858</td>
<td>-0.379</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>1.706</td>
<td>3.316</td>
<td>1.160</td>
<td>-0.749</td>
<td>-0.956</td>
<td></td>
</tr>
<tr>
<td>Boys IN athletics (C) (1)</td>
<td>1.230</td>
<td>1.507</td>
<td>0.248</td>
<td>0.102</td>
<td>-0.531</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>1.369</td>
<td>-0.693</td>
<td>-0.755</td>
<td>-0.035</td>
<td>-0.308</td>
<td></td>
</tr>
<tr>
<td>Girls NOT in athletics (C) (1)</td>
<td>11.096</td>
<td>4.902</td>
<td>1.772</td>
<td>-0.347</td>
<td>-1.263</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>11.549</td>
<td>8.802</td>
<td>2.877</td>
<td>-0.347</td>
<td>-1.263</td>
<td></td>
</tr>
<tr>
<td>Girls IN athletics (C) (2)</td>
<td>2.757</td>
<td>5.040</td>
<td>2.748</td>
<td>-0.664</td>
<td>-0.227</td>
<td></td>
</tr>
<tr>
<td>Women NOT in athletics (C) (1)</td>
<td>1.585</td>
<td>-0.351</td>
<td>-0.616</td>
<td>0.265</td>
<td>-0.470</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>1.574</td>
<td>1.129</td>
<td>0.004</td>
<td>-0.177</td>
<td>-0.893</td>
<td></td>
</tr>
<tr>
<td>Women IN athletics (B)</td>
<td>0.953</td>
<td>0.241</td>
<td>0.284</td>
<td>0.905</td>
<td>0.381</td>
<td></td>
</tr>
<tr>
<td>Participation (B)</td>
<td>0.968</td>
<td>-1.754</td>
<td>-0.503</td>
<td>-0.200</td>
<td>0.435</td>
<td></td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (1)</td>
<td>2.899</td>
<td>5.116</td>
<td>1.972</td>
<td>0.788</td>
<td>-1.519</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>2.330</td>
<td>4.225</td>
<td>1.913</td>
<td>0.788</td>
<td>-1.519</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>1.848</td>
<td>2.886</td>
<td>0.927</td>
<td>-1.116</td>
<td>-1.330</td>
<td></td>
</tr>
</tbody>
</table>

The interview with respondent 110 and the answers to the questionnaire were again examined. There were no obvious reasons for this case not being considered as part of the population. From the practical experience of the researcher, the characteristics of respondent 110 are not unique in athletics in the UK.

It was decided to leave the case in the analysis and accept the Thesis Male model as being acceptable for the purposes of this research. This decision took into account that the purpose of the research was to show that logistic regression could be used to model the coach decision by significantly fitting a set of personal and attitudinal variables.
10.7. Relative Significance of Variables

The estimated values of the coefficients do not indicate their relative effects as the scales used to measure each variable differ. For example, “Gender” is a binary variable with a scale ranging from 0 to 1. At the other extreme, “Flesch” is a continuous variable with a theoretical range from 0 to 100. As the model is estimating a binary variable, “Coach”, with a range from 0 to 1, the absolute value of each coefficient will be associated with the range of values in the data of that variable. Variables with large ranges will tend to have higher coefficients than those with lower ranges, regardless of their relative effects.

This problem can be overcome by standardising the coefficient values. The standardised logistic regression coefficient is estimated as follows:

\[ b_{yx}^* = \frac{(b_{yx})(S_x)(R)}{S_{logit}(\hat{y})} \]  

(Equation 10.7.1.)

where:

- \( b_{yx}^* \) is the standardised logistic regression coefficient for coefficient \( yx \)
- \( b_{yx} \) is the model estimated logistic regression coefficient for coefficient \( yx \)
- \( S_x \) is the standard deviation of the explanatory variable \( x \)
- \( R \) is the square root of the Pearson Correlation Coefficient of the actual to predicted values of the response variable.
- \( \hat{Y}_{yx} \) is the predicted value of the response variable for case \( y \)

\[ \text{logit}(\hat{Y}) = \ln\left(\frac{\hat{Y}}{1 - \hat{Y}}\right) \]
The standardised logistic regression coefficient of an explanatory variable is interpreted as follows. A one standard deviation change in the variable produces a $b_{YX}^*$ standard deviation change in logit ($Y$) (Menard, S, 1995).

The standardised logistic regression coefficients for the Thesis Model are shown in Table 10.7.1. in the rank order of their absolute values. The variable at the head of the Table has the strongest effect on the model. It will be seen that the strongest variables have effects of 10 or greater than the weakest variables.

### Table 10.7.1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standardised</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls NOT in athletics (C) (2)</td>
<td>-17.970</td>
<td>0.695</td>
<td>1</td>
</tr>
<tr>
<td>Girls IN athletics (C) (1)</td>
<td>-15.007</td>
<td>0.637</td>
<td>2</td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (1)</td>
<td>-13.761</td>
<td>0.636</td>
<td>3</td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C) (3)</td>
<td>13.260</td>
<td>0.575</td>
<td>4</td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C) (1)</td>
<td>13.062</td>
<td>0.530</td>
<td>5</td>
</tr>
<tr>
<td>Girls NOT in athletics (C) (1)</td>
<td>-12.481</td>
<td>0.529</td>
<td>6</td>
</tr>
<tr>
<td>Cars available (C) (1)</td>
<td>10.510</td>
<td>0.495</td>
<td>7</td>
</tr>
<tr>
<td>Minutes to track (B)</td>
<td>11.592</td>
<td>0.470</td>
<td>8</td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C) (2)</td>
<td>9.299</td>
<td>0.403</td>
<td>9</td>
</tr>
<tr>
<td>Time home (C) (1)</td>
<td>-8.735</td>
<td>0.371</td>
<td>10</td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (2)</td>
<td>-8.686</td>
<td>0.360</td>
<td>11</td>
</tr>
<tr>
<td>Time home (C) (2)</td>
<td>-6.979</td>
<td>0.329</td>
<td>12</td>
</tr>
<tr>
<td>Club (B)</td>
<td>5.077</td>
<td>0.239</td>
<td>13</td>
</tr>
<tr>
<td>Cars available (C) (2)</td>
<td>4.503</td>
<td>0.208</td>
<td>14</td>
</tr>
<tr>
<td>Participation (B)</td>
<td>4.084</td>
<td>0.173</td>
<td>15</td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (3)</td>
<td>-3.798</td>
<td>0.165</td>
<td>16</td>
</tr>
<tr>
<td>Women NOT in athletics (C) (1)</td>
<td>2.952</td>
<td>0.123</td>
<td>17</td>
</tr>
<tr>
<td>Boys IN athletics (C) (1)</td>
<td>-3.507</td>
<td>0.116</td>
<td>18</td>
</tr>
<tr>
<td>Girls IN athletics (C) (2)</td>
<td>-2.873</td>
<td>0.114</td>
<td>19</td>
</tr>
<tr>
<td>Boys IN athletics (C) (2)</td>
<td>3.231</td>
<td>0.094</td>
<td>20</td>
</tr>
<tr>
<td>Women IN athletics (B)</td>
<td>-2.880</td>
<td>0.081</td>
<td>21</td>
</tr>
<tr>
<td>Women NOT in athletics (C) (2)</td>
<td>0.912</td>
<td>0.041</td>
<td>22</td>
</tr>
</tbody>
</table>

**DATA & MODELLING**
The effects can be more clearly seen in Diagram 10.7.1. Those variables having a positive effect on the decision to become a coach are in black. Those variables having a negative effect on the decision are shaded with diagonal lines.

Consideration of standardised coefficients looks at how each variable relatively effects the value of the response variable in the model. It does not give information on how each explanatory variable contributes to the fit of the model. The fit contribution can be measured by the contribution that each explanatory variable makes to the log likelihood (-2LL) of the model. This contribution of -2LL can be found by running the model with and without each explanatory variable and measuring the change in -2LL. The results are shown in Table 10.7.2 and graphically in Diagram 10.7.2. The results are shown in the order of the first appearance of the variable in Table 10.7.1. (standardised coefficients).

It will be seen that the relative effect of a variable is not necessarily related to the contribution to fit of that variable. In particular, the variable “Girls NOT in athletics (C)” contributes relatively little to the fit, as measured by log likelihood (-2LL), but
has, relatively, the largest effect on the response variable. Conversely, the variable "Boys IN athletics (C) (1)" contribute, relatively, quite a lot to fit, but has, relatively little effect on the response variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>-2LL Contributed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls NOT in athletics (C)</td>
<td>22.908</td>
</tr>
<tr>
<td>Girls IN athletics (C)</td>
<td>134.647</td>
</tr>
<tr>
<td>Attitude; pay coaches (C)</td>
<td>87.685</td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C)</td>
<td>36.797</td>
</tr>
<tr>
<td>Cars available (C)</td>
<td>86.052</td>
</tr>
<tr>
<td>Minutes to track (B)</td>
<td>9.313</td>
</tr>
<tr>
<td>Time home (C)</td>
<td>36.841</td>
</tr>
<tr>
<td>Club (B)</td>
<td>26.987</td>
</tr>
<tr>
<td>Participation (B)</td>
<td>25.084</td>
</tr>
<tr>
<td>Women NOT in athletics (C)</td>
<td>14.944</td>
</tr>
<tr>
<td>Boys IN athletics (C)</td>
<td>50.960</td>
</tr>
<tr>
<td>Women IN athletics (B)</td>
<td>8.358</td>
</tr>
</tbody>
</table>

These variances can be more clearly seen in the graphical representation of the log likelihood contributions in Diagram 10.7.2.

It will be seen that the rank order of the contribution to fit measured by -2LL differ to the rank order of the standardised coefficients. For example, the variable "Girls NOT in athletics" has the largest standardised coefficient, indicating that this variable contributes the greatest effect on the decision process. Yet, amongst the significant variables, it makes one of the of the lowest contributions to fit, as measured by -2LL contribution. The converse applies to the variable ‘Boys IN athletics’, which has a low standardised coefficient and a, relatively, high -2LL contribution.

These apparent contradiction illustrate the complexity of a multi-dimensional process. The standardised coefficient indicates the relative importance of a single variable on the decision of an individual within the sample. The -2LL contribution is one way of indicating the predictive importance of that variable on the sample as a whole. Which to use depends on the use required.
But fit is not the only aspect of a model that may be of interest. The other aspect is the assignment ability of the model and how each variable effects this ability. logistic regression models concern response variables that are binary, in the case of the model in this research: “Coach” and “Not Coach”. The ability to assign has two components, one related to each of the binary states. Each component is measured by the percentage of cases that are correctly assigned. In the case of the research model, this is the percentage of “Not Coach” cases that are correctly assigned by the model to “Not Coach” and the percentage of “Coach” cases that are correctly assigned to “Coach”.

Diagram 10.7.2
-2LL CONTRIBUTION TO FINAL MODEL - BY STANDARDISED COEFFICIENT RANK

<table>
<thead>
<tr>
<th>EXPLANATORY VARIABLE REMOVED</th>
<th>-2LL LOST ON REMOVAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls NOT in athletics</td>
<td></td>
</tr>
<tr>
<td>Girls IN athletics</td>
<td></td>
</tr>
<tr>
<td>Attitude to paying coaches</td>
<td></td>
</tr>
<tr>
<td>Gender &amp; Regularity of work hours</td>
<td></td>
</tr>
<tr>
<td>No. of cars</td>
<td></td>
</tr>
<tr>
<td>Time to track</td>
<td></td>
</tr>
<tr>
<td>Time home</td>
<td></td>
</tr>
<tr>
<td>Club</td>
<td></td>
</tr>
<tr>
<td>Participation frequency</td>
<td></td>
</tr>
<tr>
<td>Women NOT in athletics</td>
<td></td>
</tr>
<tr>
<td>Boys IN athletics</td>
<td></td>
</tr>
<tr>
<td>Women IN athletics</td>
<td></td>
</tr>
</tbody>
</table>

DATA & MODELLING
In the Final Model, “Not Coach” cases have a 100.0% correct assignment to “Not Coach”. This 100% assignment does not change is not reduced by the removal of any one of the explanatory variables. The effect of removal of each variable on the correct “Coach” assignments is shown in Table 10.7.3. and graphically illustrated in Diagram 10.7.3. The rank order of variables for the standardised coefficients is retained to allow comparison of the assignment table and Diagram (10.7.3) with the coefficient table and Diagram (10.7.1) and the -2LL table and Diagram (10.7.2).

It will be seen that the relative contributions of each explanatory variable to correct assignment can be quite different to the effect of that variable on the response variable and upon the contribution of log likelihood to the model. In particular, “Girls NOT in athletics”, “Minutes to track”; and “Boys IN athletics” show marked differences.

<table>
<thead>
<tr>
<th>Variable</th>
<th>% Correct</th>
<th>Variable removed</th>
<th>Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls NOT in athletics (C)</td>
<td>78.8%</td>
<td>74.2%</td>
<td>3.6%</td>
</tr>
<tr>
<td>Girls IN athletics (C)</td>
<td>78.8%</td>
<td>43.0%</td>
<td>34.5%</td>
</tr>
<tr>
<td>Attitude; pay coaches (C)</td>
<td>78.8%</td>
<td>48.7%</td>
<td>29.1%</td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C)</td>
<td>78.8%</td>
<td>66.5%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Cars available (C)</td>
<td>78.8%</td>
<td>54.7%</td>
<td>23.6%</td>
</tr>
<tr>
<td>Minutes to track (B)</td>
<td>78.8%</td>
<td>73.5%</td>
<td>7.6%</td>
</tr>
<tr>
<td>Time home (C)</td>
<td>78.8%</td>
<td>62.7%</td>
<td>16.4%</td>
</tr>
<tr>
<td>Club (B)</td>
<td>78.8%</td>
<td>69.3%</td>
<td>9.1%</td>
</tr>
<tr>
<td>Participation (B)</td>
<td>78.8%</td>
<td>60.1%</td>
<td>18.2%</td>
</tr>
<tr>
<td>Women NOT in athletics (C)</td>
<td>78.8%</td>
<td>69.0%</td>
<td>9.1%</td>
</tr>
<tr>
<td>Boys IN athletics (C)</td>
<td>78.8%</td>
<td>63.9%</td>
<td>14.5%</td>
</tr>
<tr>
<td>Women IN athletics (B)</td>
<td>78.8%</td>
<td>72.3%</td>
<td>5.4%</td>
</tr>
</tbody>
</table>
Diagram 10.7.3.

REDUCTION IN % OF “Coach” CORRECTLY ASSIGNED BY VARIABLE

The following Section considers the regression coefficients and how they may be interpreted.


Whilst the relative effects of the explanatory variables were discussed in Section 10.6., the calculated values of the regression coefficients must be considered.

In linear regression, the value of the regression coefficient represents the amount by which the response variable increases with each one unit increase in the explanatory variable.
The response variable is measured in the same units as it was observed in the real world. If the explanatory variable is not derived from a transformation of the observed data, the explanatory variable is also measured in the same units as that variable was measured in the real world. The regression coefficient directly measures the effect of the explanatory variable on the response variable. For example, there is a relationship between body weight (measured in pounds) and an individual’s strength (measured in units of the test machine). The relationship can be expressed as a linear regression of the form: \( \text{Strength} = A \text{constant} + A \text{coefficient} \times \text{Body Weight} \). The constant has been found to be -66.24 and the coefficient 2.38. The relationship is such that strength increases at the rate of 2.38 units of the test machine for every 1 pound of body weight.

In logistic regression, the response variable is measured in a transform of the units in it was observed. In this research, the response variable was observed as a binary state of the respondent being or not being a coach. The transformation is the natural logarithm of the calculated odds ratio of being a coach, \( \logit(Y) \). The odds ratio is the calculated by dividing the probability of being a coach by 1 - the probability of being a coach (see Section 7.1.). What is important is that the calculated coefficients represent the change in \( \logit(Y) \) of the response variable for a one unit change in the explanatory variable.

Where they are not transformed, the explanatory variables are measured in the same units as the observed values (assuming no transformations). In this research, all the explanatory values in the Thesis model have been transformed from their original observed units. All of the model explanatory variables have been transformed into categorical variables. The method of using categorical explanatory variables in logistic regression is to represent them as sets of binary states. Then each category in a set is compared to a reference category (see Section 7.6.). In this research, the reference category is always the last category. In the case of a categorical explanatory variable, the coefficient represents the change in \( \logit(Y) \) of the response.
variable when the category of the explanatory variable changes from the reference category, the last, to the category of the categorical explanatory variable.

For example, the logistic regression coefficient in the Thesis Model for the explanatory variable “attitude to paying coaches (3)” is -3.798 (see Table 10.5.2.). Category (3) of this variable are the responses of “Neutral & Don’t know” to the statement that “Coaches at local sports clubs should generally be paid”. The reference category for the same variable are the response of “Disagree” to the same statement. So the coefficient of the variable indicates that logit (Y) changes by -13.761 when the category is “Neutral” in comparison to when the category is “Disagree”. The negative sign indicates that the logit decreases. This means that the probability of being a coach decreases.

As the transformation involves taking the natural logarithm of the odds ratio, it is possible to transform each coefficient by taking the exponential of its calculated value. These values are referred to as exp(B) and are in units of odds ratios. The odds ratio represents the change in the odds for the response variable for each one unit change in the explanatory variable.

Following the example above of the explanatory variable “attitude to paying coaches (3)”, the exponential of -3.798 is 0.022 (exp(B)). This can be interpreted as the odds of being a coach diminish by 97.8% when a respondent has a neutral attitude to paying coaches compared to an attitude of agreement.

The calculated coefficients, their exp(B) values and the confidence intervals for exp(B) (confidence level of 95%) are shown in Table 10.8.1. It will be seen that many of the values of exp(B) and their confidence intervals are either very low (less than 0.001) or are very high (greater than 1,000). These extremes arise from the fact that the calculated values of the coefficients are either positively or negatively very high. The large coefficients arise from the size of sample (n=112) in relation to the number of variables and categories (22) and the overdispersion (see 10.6.1.).
Table 10.8.1. 

COEFFICIENTS, Exp(B) AND CONFIDENCE LIMITS - THESIS MODEL

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Exp(B)</th>
<th>Mean</th>
<th>Upper Limit</th>
<th>Lower Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls NOT in athletics (C) (2)</td>
<td>-17.970</td>
<td>very small</td>
<td>very small</td>
<td>106.146</td>
<td></td>
</tr>
<tr>
<td>Girls IN athletics (C) (1)</td>
<td>-15.007</td>
<td>very small</td>
<td>very small</td>
<td>very small</td>
<td></td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (1)</td>
<td>-13.761</td>
<td>very small</td>
<td>very small</td>
<td>very small</td>
<td></td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C) (3)</td>
<td>13.260</td>
<td>very large</td>
<td>887.393</td>
<td>very large</td>
<td></td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C) (1)</td>
<td>13.062</td>
<td>very large</td>
<td>645.527</td>
<td>very large</td>
<td></td>
</tr>
<tr>
<td>Girls NOT in athletics (C) (1)</td>
<td>-12.481</td>
<td>very small</td>
<td>very small</td>
<td>very large</td>
<td></td>
</tr>
<tr>
<td>Cars available (C) (1)</td>
<td>10.510</td>
<td>very small</td>
<td>396.301</td>
<td>very large</td>
<td></td>
</tr>
<tr>
<td>Minutes to track (B)</td>
<td>11.592</td>
<td>very large</td>
<td>16.869</td>
<td>very large</td>
<td></td>
</tr>
<tr>
<td>Gender &amp; Regular Hrs (C) (2)</td>
<td>9.299</td>
<td>very small</td>
<td>81.405</td>
<td>very large</td>
<td></td>
</tr>
<tr>
<td>Time home (C) (1)</td>
<td>-8.735</td>
<td>0.030</td>
<td>very small</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (2)</td>
<td>-8.686</td>
<td>very small</td>
<td>very small</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>Time home (C) (2)</td>
<td>-6.979</td>
<td>very small</td>
<td>very small</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td>Club (B)</td>
<td>5.077</td>
<td>160.354</td>
<td>16.869</td>
<td>very large</td>
<td></td>
</tr>
<tr>
<td>Cars available (C) (2)</td>
<td>4.503</td>
<td>90.257</td>
<td>2.961</td>
<td>very large</td>
<td></td>
</tr>
<tr>
<td>Participation (B)</td>
<td>4.084</td>
<td>59.406</td>
<td>8.918</td>
<td>395.732</td>
<td></td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (3)</td>
<td>-3.798</td>
<td>0.022</td>
<td>0.001</td>
<td>0.838</td>
<td></td>
</tr>
<tr>
<td>Women NOT in athletics (C) (1)</td>
<td>2.952</td>
<td>very small</td>
<td>0.635</td>
<td>316.696</td>
<td></td>
</tr>
<tr>
<td>Boys IN athletics (C) (1)</td>
<td>-3.507</td>
<td>0.030</td>
<td>0.003</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td>Girls IN athletics (C) (2)</td>
<td>-2.873</td>
<td>very small</td>
<td>0.003</td>
<td>0.962</td>
<td></td>
</tr>
<tr>
<td>Boys IN athletics (C) (2)</td>
<td>2.321</td>
<td>25.046</td>
<td>1.712</td>
<td>366.323</td>
<td></td>
</tr>
<tr>
<td>Women IN athletics (B)</td>
<td>-2.880</td>
<td>0.056</td>
<td>0.009</td>
<td>0.363</td>
<td></td>
</tr>
<tr>
<td>Women NOT in athletics (C) (2)</td>
<td>0.912</td>
<td>0.402</td>
<td>0.018</td>
<td>8.780</td>
<td></td>
</tr>
</tbody>
</table>

Very large: >1,000,000  Very small: <0.001

The results are unsatisfactory for model interpretation because of the large number of variables having either very high or very low values and the wide confidence intervals for every variable. However, the relative influence of each variable can be observed by the calculation of the standardised coefficients reported in Section 10.7.1. It should be noted that the rank order of variables by exp(B) is the same as the rank order by standardised regression coefficient. The latter is the order of variables in Table 10.8.1. above.
Some idea of the effects of the more important variables can be gained by building a Reduced Thesis Model using the 5 explanatory variables that add the greatest amounts of -2LL to the model. This method increases the ratio of number of cases to number of variables giving tighter confidence limits to the coefficients. The results are shown in Table 10.8.2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Exp(B)</th>
<th>Mean</th>
<th>Upper Limit</th>
<th>Lower Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude; pay coaches (C) (1)</td>
<td>-5.828</td>
<td>0.003</td>
<td>0.023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (2)</td>
<td>-3.209</td>
<td>0.040</td>
<td>0.213</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (3)</td>
<td>-0.622</td>
<td>0.537</td>
<td>2.609</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls IN athletics (C) (1)</td>
<td>-4.582</td>
<td>0.010</td>
<td>0.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls IN athletics (C) (2)</td>
<td>-0.894</td>
<td>0.409</td>
<td>1.689</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cars available (C) (1)</td>
<td>6.411</td>
<td>608.359</td>
<td>7016.284</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cars available (C) (2)</td>
<td>3.378</td>
<td>29.304</td>
<td>239.696</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys IN athletics (C) (1)</td>
<td>-1.252</td>
<td>0.286</td>
<td>1.320</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys IN athletics (C) (2)</td>
<td>2.049</td>
<td>7.763</td>
<td>43.021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time home (C) (1)</td>
<td>-2.755</td>
<td>0.064</td>
<td>0.222</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time home (C) (2)</td>
<td>-2.371</td>
<td>0.093</td>
<td>0.257</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table 10.8.2., three variables have values of exp(B) that may be zero. The indication that the value may be zero is that the upper and lower limits span the value of zero (shaded cells to right). Each of these model variables is a category within one of the categorical variables. If the true value of exp(B) is zero, then that category can be merged with one of the other categories within that categorical variable. In addition, the two coefficients for the variable “Time home” are, by comparison with their standard errors, probably not significantly different.

These four variables are recoded as follows:

Attitude to paying coaches (C) (1) = 1 when Attitude is “Strongly agree” or “Agree”

Attitude to paying coaches (C) (2) = 2 when Attitude is “Neutral or Don’t know”
Attitude to paying coaches (C) (2) = 3 when Attitude is “Strongly Disagree” or “Disagree”

Girls IN athletics (B) (1) = 0 when number of Girls IN athletics = 0
Girls IN athletics (B) (2) = 1 when number of Girls IN athletics > 0

Boys IN athletics (B) (1) = 0 when number of Boys IN athletics = 0
Boys IN athletics (B) (2) = 1 when number of Boys IN athletics > 0

Free (B) (1) = 0 when home by 6 p.m.
Free (B) (2) = 1 when home after 6 p.m.

The results are shown in Table 10.8.3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Value</th>
<th>Stand. Error</th>
<th>Mean</th>
<th>Upper Limit</th>
<th>Lower Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude; pay coaches (C) (1)</td>
<td></td>
<td>-4.489</td>
<td>0.671</td>
<td>0.011</td>
<td>0.003</td>
<td>0.042</td>
</tr>
<tr>
<td>Attitude; pay coaches (C) (2)</td>
<td></td>
<td>-1.750</td>
<td>0.415</td>
<td>0.174</td>
<td>0.077</td>
<td>0.392</td>
</tr>
<tr>
<td>Girls IN athletics (B) (1)</td>
<td></td>
<td>-3.205</td>
<td>0.438</td>
<td>0.041</td>
<td>0.017</td>
<td>0.096</td>
</tr>
<tr>
<td>Cars available (C) (1)</td>
<td></td>
<td>4.625</td>
<td>0.754</td>
<td>101.990</td>
<td>23.255</td>
<td>447.298</td>
</tr>
<tr>
<td>Cars available (C) (2)</td>
<td></td>
<td>1.358</td>
<td>0.448</td>
<td>3.889</td>
<td>1.615</td>
<td>9.364</td>
</tr>
<tr>
<td>Boys IN athletics (B) (1)</td>
<td></td>
<td>-2.220</td>
<td>0.398</td>
<td>0.109</td>
<td>0.050</td>
<td>0.237</td>
</tr>
<tr>
<td>Time home (B) (1)</td>
<td></td>
<td>-1.710</td>
<td>0.379</td>
<td>0.181</td>
<td>0.086</td>
<td>0.380</td>
</tr>
</tbody>
</table>

It can be seen from Table 10.8.3, all the coefficients are reasonable and, categorical, consistent; the standard errors are acceptable; and the values of $\exp(B)$ interpretable.

The Thesis Model is interpreted in the next Section, 10.9, by comparing the results with the original hypotheses in Section 9.2, explaining apparent anomalies. In the following Section, 10.9, the direction of the effect of each variable is considered by comparison with logical expectations.
In this Section the hypotheses put forward in Section 9.2. for each of the potential explanatory variables are compared to the presence or absence in the Thesis model and to the sign of their coefficients in the Thesis Model. In some instances, the explanatory variable is in the model and the direction of the coefficient conforms to the hypothesis. In other instances, the explanatory variable is in the model, but the direction of the coefficient contradicts the hypothesis. In yet other instances, the explanatory variable is not in the model. Absence from the Thesis model is either because the variable was found to be highly correlated with a variable in the model or because the variable did not contribute significantly to the fit of the model. This stage of the interpretation is summarised in Table 10.9.1.

It will be seen that some variables and variable groups in the Thesis Model have directions of effect opposite to those hypothesised and one variable group, Dependants, has some categories with the same direction of effect as the hypothesis and some variables in the opposite direction. It will also be seen that several hypothesised variables and variable groups could not be included in the model as they were correlated with variables in the Thesis Model. “Positive” indicates that being in the category compared to being in the reference category of the variable is associated with an increase in the odds of being a coach. A “Negative” coefficient is associated with a decrease in the odds.

Possible reasons for directions differing from the hypotheses are discussed after the Table. Following this, the correlations are discussed. Possible reasons are discussed in Sections 10.9.1. and 10.9.2. for the two explanatory variables that exhibit differences in direction of the coefficients between the original hypotheses and the model.
### Table 10.9.1.
COMPARISON OF HYPOTHESES WITH THESIS MODEL

<table>
<thead>
<tr>
<th>Variable Group</th>
<th>In either Model?</th>
<th>If not in Model, Reason</th>
<th>Direction of Effect Hypothesis</th>
<th>Direction of Effect Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (compared to female)</td>
<td>Yes, as Interaction</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>Dependants (compared to several)</td>
<td>Five variables remaining 3 not significant</td>
<td>Decreasing</td>
<td>Variable</td>
<td></td>
</tr>
<tr>
<td>Distance to track (compared to far)</td>
<td>Yes</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>No. of cars (compared to 3+)</td>
<td>Yes</td>
<td>Increasing</td>
<td>Decreasing</td>
<td></td>
</tr>
<tr>
<td>Education: School leaving age, Years of Post School Education</td>
<td>No Correlated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time home (compared to after 6 p.m.)</td>
<td>Yes</td>
<td>Decreasing</td>
<td>Decreasing</td>
<td></td>
</tr>
<tr>
<td>Regular hours (compared to irregular)</td>
<td>Yes, as Interaction</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>Other sports (compared to more than athletics)</td>
<td>No Not significant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years competing (compared to &gt;50%)</td>
<td>No Correlated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No interest in Veteran athletics (compared to interest)</td>
<td>No Not significant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No interest in Road Running (compared to interest)</td>
<td>No Not significant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participate 1/month or more (compared to less)</td>
<td>Yes</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>&gt;50% of work is sport (compared to less)</td>
<td>No Correlated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literacy: Word complexity, Sentence length, Flesch score.</td>
<td>No Not significant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Don’t read Science (compared to read)</td>
<td>No Correlated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disagree on others should volunteer (compare to Agree)</td>
<td>No Correlated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;49% Work involves sport (compare &lt;50%)</td>
<td>No Not significant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly Agree on paying coaches (compared Disagree)</td>
<td>Yes Increasing</td>
<td>Increasing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at decision</td>
<td>No Correlated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year of birth: Generation</td>
<td>No Correlated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Club A (compared Club B)</td>
<td>Yes No hypothesis</td>
<td>Positive</td>
<td>Positive</td>
<td></td>
</tr>
</tbody>
</table>
Following these Sections, Section 10.10. discusses the interpretation of the signs of the coefficients and Section 10.11. discusses the explanatory variables not included in the Thesis Model due to correlations.

10.9.1. DEPENDANTS

The original hypothesis for dependants was that the more there were, the lower the probability of the respondent being a coach. In terms of the model in which all variables are categorical with the last category (the highest number of dependants) as the reference category, the hypothesis would predicate that the coefficient would decrease in value as the category got nearer to the reference category, i.e. the probability of being a coach decreases as the number of dependants increases (the reference category having the largest number). Five explanatory variables related to Dependants are in the Thesis Model. All were represented as categorical variables, the reference category being the last. Their coefficients, standard errors and reference categories are shown in Table 10.9.2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys IN athletics</td>
<td>(1) = None</td>
<td>-3.507</td>
<td>1.230</td>
</tr>
<tr>
<td></td>
<td>(2) = One</td>
<td>+3.221</td>
<td>1.369</td>
</tr>
<tr>
<td></td>
<td>Increasing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls NOT in athletics</td>
<td>(1) = None</td>
<td>-12.481</td>
<td>11.096</td>
</tr>
<tr>
<td></td>
<td>(2) = One</td>
<td>-17.970</td>
<td>11.549</td>
</tr>
<tr>
<td></td>
<td>Decreasing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women NOT in athletics</td>
<td>(1) = None</td>
<td>-15.007</td>
<td>2.757</td>
</tr>
<tr>
<td></td>
<td>(2) = One</td>
<td>-2.873</td>
<td>1.446</td>
</tr>
<tr>
<td></td>
<td>Increasing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women IN athletics</td>
<td>(1) = None</td>
<td>+2.652</td>
<td>1.585</td>
</tr>
<tr>
<td></td>
<td>(2) = One</td>
<td>-0.912</td>
<td>1.574</td>
</tr>
<tr>
<td></td>
<td>Decreasing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Both the categorical variables relating to dependants NOT in athletics, Girls NOT in athletics and Women NOT in athletics, have values of the coefficients decreasing from

**DATA & MODELLING**
category (1) to category (2), i.e. the odds of being a coach decrease with the number of dependants. This confirms the original hypothesis.

Both the categorical variables relating to dependants IN athletics, Boys IN athletics and Girls IN athletics, have values of the coefficients increasing from category (1) to category (2), i.e. the odds of being a coach increase with the number of dependants. This is contrary to the original hypothesis. But, logically, it is plausible. The more dependants, especially children, who are involved in athletics, the greater the incentive for the respondent to become a coach, especially if the respondent is a parent.

11.9.2. CARS AVAILABLE

The original hypothesis associated with number of cars available is that the greater the number, the greater the probability of being a coach. The construct behind this hypothesis is that the greater the number of cars, the less competition there is in the household for transport. The hypothesis predicates that the coefficient will increase with the number of cars. In terms of the Thesis Model, the categorical variable for cars increases as the category approaches the reference category (3), “More than 2 cars”. The coefficients in the model decrease as they reach the reference category, contrary to the original hypothesis. Having no car or one car available is associated with a large increase in the odds of being a coach compared to having 3 or more cars. The increase in odds is smaller when comparing having 2 cars with having 3 or more. Calculation of correlation coefficients of variables with “Number of cars available” and inspection of associated scatter diagrams, to reject false relationships, reveals five variables which have significant, but weak, correlations and acceptable relationships with “Number of cars available”. Pearson's Product Moment Correlation Coefficient (PR²) was used when both variables were continuous and Spearman's rho (SR²) used when either or both variables were categorical (Black, T R, 1999).
i. Time home and free, which has a negative association \( \text{PRZ} = -0.348 \) i.e. the earlier home the dependent is, the more cars available.

ii. Time from home to track, which has a negative association \( \text{PRZ} = -0.250 \)
i.e. the further away a respondent lives, the fewer cars are available.

iii. Household involved in other sports, which has a positive association \( \text{PRZ} = 0.327 \) i.e. households involved in other sports, compared to households only involved in athletics, the more cars available.

iv. Respondent interested in veterans athletics, which has a negative association \( \text{SR2} = -0.344 \) i.e. an interest in veterans athletics, as compared to not interested, the fewer cars available.

v. Number of dependants, which has a positive association \( \text{PRZ} = 0.513 \) i.e. the larger the number of dependants, the more cars available.

Variables (i), Hour home, and (ii), Time to track, are in the Thesis Model. The correlations with Number of cars are relatively weak and are insufficient to produce Tolerance statistics below 0.20. The inclusion in the Model with the variable Number of cars is unlikely to produce major effects on the Model.

Variables (iii), Household involved in other sports, and (v), Number of dependants, are significantly and positively correlated with each other \( \text{SR2} = 0.529 \), i.e. larger households tend to be more associated with other sports than smaller households. This suggests that either they need more transport to engage in more sports or that greater transport availability may allow a wider range of sport activities. As the number of sports increases, the dependent has a greater range of sports to take an interest in, including coaching. In other words, there is competition for the leisure time of the respondent. Another interpretation is that number of cars may be a surrogate measure of socio-economic group. This interpretation could accord with...
the observation in the Scottish survey of coaches that athletics coaches tended to leave school earlier than coaches of the other sports surveyed ((Lyle, J, Allison, M, and Taylor, J, 1997). This suggests that the distribution over the socio-economic groups of athletics coaches may be more frequent in the lower socio-economic groups than the distribution for non coaches.

Variable (v), respondent interested in veterans athletics, is negatively correlated with age ($\text{SR}^2 = -0.203$), i.e. there is a weak association between no interest in veteran athletics and older age groups. This can be explained for two reasons. Veterans athletics is, relatively, a new branch of the sport and, therefore, was not available to older respondents. In addition, younger respondents may see veterans athletics as being one way of extending their competitive participation in sport. In Section 10.9, it will be seen that many younger respondents who are actively competing in athletics state that they intend to continue competing as veterans and that this will take priority over them becoming a coach.

These interpretations can account for the apparent anomalies between the direction of effect in the original hypotheses and the direction of effect in the model.

10.10. Coefficient Signs

As the model is built without a constant, i.e. that without the influence of any explanatory variables the odds of becoming a coach are 0.5 (the logit is 0.0 or Not a coach), the sign of the coefficient of an explanatory variable represents whether the variable represents a relative incentive to be a coach or a relative disincentive to be a coach, the relativity being in relation to the overall effect of all the other variables in the model. The equivalent signs are positive and negative respectively.

Table 10.10.1. shows the hypothesised or expected states compared to the Thesis Model results, for incentive or disincentive, coefficient sign and direction of the coefficient values for each explanatory variable in the Thesis Model. In the case of
categorical variables, the Table shows whether or not coefficient values for the classes move towards zero (declining) or away from zero (increasing).

Disagreements between the Thesis Model and the hypothesis or expected are indicated by emboldening. It will be seen that with most variables there is agreement between the hypothesised or expected and the Thesis Model. The two exceptions are “Number of Cars available” and “Hour Home”.

Table 10.10.1.
COMPARISON OF INCENTIVE/DISINCENTIVE EXPECTATIONS WITH THESIS MODEL

<table>
<thead>
<tr>
<th>Variable Group</th>
<th>Incentive or Expected</th>
<th>Disincentive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (compared to female)</td>
<td>Incentive</td>
<td>Incentive</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>No Child Dependents or 1 (compared to 2 or more)</td>
<td>Disincentive</td>
<td>Disincentive</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>Increasing</td>
<td>Increasing</td>
</tr>
<tr>
<td>No Adult Dependents or 1 (compared to 2 or more)</td>
<td>Disincentive</td>
<td>Variable</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increasing</td>
<td></td>
</tr>
<tr>
<td>0 &amp; 1 or 2 cars available (compared to 3 or more)</td>
<td>Incentive</td>
<td>Disincentive</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>Declining</td>
<td>Declining</td>
</tr>
<tr>
<td>Home by 6 p.m. or by 7 p.m. (compared to after 8 p.m.)</td>
<td>Incentive</td>
<td>Disincentive</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>Declining</td>
<td>Declining</td>
</tr>
<tr>
<td>Regular hours (compared to irregular)</td>
<td>Incentive</td>
<td>Incentive</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Participate 1/month or more (compared to less than 1)</td>
<td>Incentive</td>
<td>Incentive</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Strongly Agree or Neutral on paying coaches (compared to Disagree)</td>
<td>Disincentive</td>
<td>Disincentive</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>Declining</td>
<td>Declining</td>
</tr>
<tr>
<td>Club A (compared Club B)</td>
<td>No hypothesis</td>
<td>Positive</td>
</tr>
</tbody>
</table>

The apparent discrepancies arising with the explanatory variables “Number of Cars available” and “Hour Home” were discussed previously in Section 10.9.2. In

The interpretation is extended in Section 10.11. to consideration of the potential explanatory variables that are not in the Thesis Model.
10.11. Variables Not in the Model

There are three main reasons why additional explanatory variables are not added to the Thesis Model:

i. The variable to be added is correlated with variables already in the Model to the extent that one or more values of the Tolerance statistic fall below 0.200.

ii. The increase in log likelihood, -2LL, by the addition of the variable is not significant, i.e. the significance of the additional -2LL is greater than 0.010.

iii. The addition of the variable causes computational problems as a result of the ratio of number of cases to number of variables being too low.

In the case of reason (i), it was found that either the added variable had a Tolerance statistic below 0.200 or its addition caused the variable “Attitude to paying coaches” to fall below its model value of 0.249. In no case did reason (iii), computational failure, occur.

Table 10.11.1 summarises the reasons for each variable being excluded from the Thesis Model. This gives the Tolerance statistic for each variable when it is entered into the Thesis Model, together with the resultant Tolerance statistic for the variable “Attitude to paying coaches” (Pay), and the significance of the additional -2LL from addition of the variable. The last column gives the reasons for excluding the variable.

The Table shows that the original hypotheses for 12 potential explanatory variables were not upheld by the sample data as the significance of their addition was greater than 0.010. Each group of these variables is considered in Sections 10.10.1. to 10.10.6. Unacceptable Tolerance statistics arise from there being high correlation
between a variable and one of more variables in the Thesis Model. 6 variables were rejected from the model for this reason. They are considered in Sections 10.11.7 to 10.11.12.

Table 10.11.1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Tolerance</th>
<th>Statistic Pay Coach</th>
<th>Significance if entered in Model?</th>
<th>Reason(s) For exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thesis Model variables</td>
<td>0.705</td>
<td>0.249</td>
<td>0.651</td>
<td>Significance</td>
</tr>
<tr>
<td>Dependants: No Boys NOT in athletics</td>
<td>0.486</td>
<td>0.249</td>
<td>0.178</td>
<td>Significance</td>
</tr>
<tr>
<td>(compared to one or more)</td>
<td>0.779</td>
<td>0.248</td>
<td>0.088</td>
<td>Significance</td>
</tr>
<tr>
<td>Dependants: No Men NOT in athletics</td>
<td>0.180</td>
<td>0.175</td>
<td>0.079</td>
<td>Variable, low Tolerance</td>
</tr>
<tr>
<td>(compared to one or more)</td>
<td></td>
<td></td>
<td></td>
<td>Significance</td>
</tr>
<tr>
<td>Dependants: No Men IN athletics</td>
<td>0.520</td>
<td>0.165</td>
<td>0.040</td>
<td>Significance</td>
</tr>
<tr>
<td>(compare after 16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education: Left school by 16</td>
<td>0.348</td>
<td>0.211</td>
<td>0.093</td>
<td>Significance</td>
</tr>
<tr>
<td>(compared to &gt;4 years)</td>
<td>0.395</td>
<td>0.184</td>
<td>0.000</td>
<td>Pay, Tolerance decrease</td>
</tr>
<tr>
<td>Other sports</td>
<td>0.273</td>
<td>0.243</td>
<td>0.755</td>
<td>Significance</td>
</tr>
<tr>
<td>(compared to more than athletics)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years competing: &lt;51%</td>
<td>0.487</td>
<td>0.248</td>
<td>0.396</td>
<td>Significance</td>
</tr>
<tr>
<td>(compared to &gt;50%)</td>
<td>0.680</td>
<td>0.245</td>
<td>0.276</td>
<td>Significance</td>
</tr>
<tr>
<td>No interest in Veteran athletics</td>
<td>0.061</td>
<td>0.138</td>
<td>0.000</td>
<td>Variable, low Tolerance</td>
</tr>
<tr>
<td>(compared to interest)</td>
<td>0.062</td>
<td>0.135</td>
<td>0.091</td>
<td>Variable, low Tolerance</td>
</tr>
<tr>
<td>&gt;50% of work is sport</td>
<td>0.076</td>
<td>0.165</td>
<td>0.177</td>
<td>Variable, low Tolerance</td>
</tr>
<tr>
<td>(compared to less)</td>
<td></td>
<td></td>
<td></td>
<td>Significance</td>
</tr>
<tr>
<td>Literacy: Word complexity</td>
<td>0.336</td>
<td>0.211</td>
<td>0.606</td>
<td>Significance</td>
</tr>
<tr>
<td>Literacy: Sentence length</td>
<td>0.110</td>
<td>0.161</td>
<td>0.000</td>
<td>Variable, low Tolerance</td>
</tr>
<tr>
<td>Literacy: Flesch Score</td>
<td>0.680</td>
<td>0.245</td>
<td>0.276</td>
<td>Significance</td>
</tr>
<tr>
<td>Don't read Science</td>
<td>0.081</td>
<td>0.158</td>
<td>0.000</td>
<td>Variable, low Tolerance</td>
</tr>
<tr>
<td>(compared to read)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disagree on others should volunteer</td>
<td>0.061</td>
<td>0.141</td>
<td>0.000</td>
<td>Variable, low Tolerance</td>
</tr>
<tr>
<td>(compare to Agree)</td>
<td>0.189</td>
<td>0.178</td>
<td>0.000</td>
<td>Variable, low Tolerance</td>
</tr>
<tr>
<td>&gt;49% Work involves sport</td>
<td>0.110</td>
<td>0.161</td>
<td>0.000</td>
<td>Variable, low Tolerance</td>
</tr>
<tr>
<td>(compared to &lt;50%)</td>
<td>0.680</td>
<td>0.245</td>
<td>0.276</td>
<td>Significance</td>
</tr>
<tr>
<td>Age at decision</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Born</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender: Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(compared to Female)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DATA & MODELLING
The variables “Boys IN athletics”, “Girls NOT in athletics”, “Women NOT in athletics” and “Women IN athletics” were in the Thesis Model. But the variables “Boys NOT in athletics”, “Men NOT in athletics” and “Men IN athletics” were not in the Thesis Model. The latter were all left out of the Thesis Model because the added -2LL resulting from their inclusion were not significant.

Inspection of the logits for each of the 8 explanatory variables listed above does not reveal any obvious differences between those from the 5 variables included and the 3 variables excluded. However inspection of the correlation coefficients with the variables and the variable “Gender” indicated some association.

The logits for each of the categories for each of the 8 variables transformed into categorical variables indicated some differences between the logits for variables in the model compared to variables not in the model. In general terms, those variables in the model had higher absolute values of the logits than those not in the model when divided into data from male and from female respondents.

This would suggest that variables with low effects, low absolute values of the logit, are not significant when they are introduced into the Model. Also, the absolute values of the logits from data from female respondents were larger than the values from male respondents. This is to be expected from the correlation with gender.

Both these observations can be seen in the graphical plots of the logits in Diagrams 10.11.1 and 10.11.2. The graphs for variables in the model are in Diagram 10.10.1.1. Note how the logits derived from data from male respondents lie higher on each graph than the logits derived from data from female respondents, indicating that the effect of male respondents is greater than the effect of female respondents. Note also that the logits from male respondent data are similar between the values for the number of dependants in each graph. This indicates that the effect of increasing the number of
dependants from male respondents is not great. Note also how the trend of the logits from male respondents reflects the trend from female respondents.

The logits for variables not in the model are in Diagram 10.11.2. Again, the logits derived from data from male respondents lie higher in each graph than those derived from female respondents and have smaller absolute values. Compare graphs in Diagrams 10.11.1. and 10.11.2. Except for the graph for Men in athletics, all the graphs for variables not in the model have lower absolute values of the logits than the
graphs for variables in the model, indicating the low effect that these variables have on
the model.

**Diagram 10.11.2.**

LOGITS OF VARIABLES RELATED TO DEPENDANTS NOT INCLUDED IN THE THESIS MODEL

These effects can be seen from the summary of logits in Table 10.11.2. Looking at
the right hand column for the range, both for male and female respondents, the logits
for both categories for the variables in the model are higher than those for variables
not in the model. Comparing the logit absolute values for male and for female
respondents for the variables in the model to those not in the model, they are greater
for those in the model.

**Table 10.11.2.**

AVERAGE LOGITS, BY IN OR OUT OF MODEL & BY GENDER OF RESPONDENT

<table>
<thead>
<tr>
<th>Variable Category</th>
<th>Respondents</th>
<th>Male</th>
<th>Female</th>
<th>Male - Female</th>
<th>Ratio</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>In Model</td>
<td>0</td>
<td>-0.36</td>
<td>-2.83</td>
<td>7.86</td>
<td>2.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.22</td>
<td>-2.75</td>
<td>12.50</td>
<td>2.97</td>
<td></td>
</tr>
<tr>
<td>Not in Model</td>
<td>0</td>
<td>-0.10</td>
<td>-1.12</td>
<td>11.20</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>-0.17</td>
<td>-1.61</td>
<td>9.47</td>
<td>1.45</td>
<td></td>
</tr>
</tbody>
</table>
What Table 10.11.2. also shows is that there is a strong association with gender of respondent and the logits with data from female respondents having greater effects, greater absolute values of the logit, than from male respondents.

This gender effect can be accommodated in logistic regression modelling in one of three ways:

i. Gender can be added to the model as a binary variable.

ii. An Interaction term between gender and the dependent variables can be put in the model.

iii. Two models can be constructed, one for data from male respondents and one from female respondents.

Addition of Gender as a variable, (i), and as an Interaction term, (ii), are not feasible as they cause Tolerance statistics to fall below 0.20 indicating considerable correlation with other variables. Construction of two models, (iii), is feasible and is reported in Chapter 11.

The effect in the Thesis Model of not making gender distinctions for the dependent variables may not be great as the ratios of the absolute values of the logits for females are about 10 times for males and are similar between categories. This is indicated in the column labelled “ratio” in Table 10.11.2. Thus the values of the coefficients in the Thesis Model will not be greatly effected by the absence of distinction, although the standard errors and hence the values of exp(B) will be enlarged.

10.11.2. REJECTED VARIABLES - EDUCATION

Both variables in this group, “Age left School” and “Years of Post School formal education”, were omitted from the Thesis Model because their addition, either singularly or in combination as “Years of formal Education since 15”, do not add a significant amount of -2LL to the Model.
The educational variables are all significantly correlated with the variable in the Model "Attitude to paying coaches". As will be seen in Section 10.11.7., they are also correlated with the variables concerning literacy. As both education and literacy are more easily interpretable than "attitude to paying coaches", tests were made to see whether the Thesis Model could be improved by substituting "attitude to paying coaches" by any or a combination of either any of the variables associated with education or with any combination of the variables associated with literacy. No such combination exists.

10.11.3. REJECTED VARIABLES - ASSOCIATION WITH SPORT

Two variables in this group are included in the Thesis Model: "Involvement of dependants with sports other than athletics" and "Current frequency of involvement with sport".

The four other variables in this group ("% of Post School years competing in any sport", "Interest in Veterans athletics", "Interest in road running" and "% of work time involving sport") do not contribute significantly to the Thesis Model if they are included.

Also the variable "% of Post School years competing in any sport" causes some of the model Tolerance statistics to fall below 0.20, indicating collinearity. Inspection of the correlation matrix indicates that this is probably because the variable is highly correlated, negatively, with some of the Dependants variables and with the variable of car availability and, positively, with the variable "attitude to paying coaches".

10.11.4. REJECTED VARIABLES - LITERACY

None of the variables in this group are in the Thesis Model. Three of the four variables associated with literacy ("Sentence length", "Flesch Ease of Reading Score" and "Read Science") do not significantly improve the Thesis Model.
complexity does add significantly, but causes some Tolerance statistics to fall below 0.20, indicating association with one or more variables in the Thesis Model.

The highest and significant correlation with a variable in the model is “Time to reach track”, which is positive, i.e. there is an association between being further from the track and reading literature with more complex words. In turn, “Word complexity” is positively associated with school leaving age. But the association between “Time to reach track” and “School leaving age” is weak, but still significant.

One possible explanation is that athletic tracks are usually situated inside towns and not in the country. It may be that, generally, people who live in the country are often more wealthy than those living in the towns. This is true for the County of Surrey (Central Statistical Office. 1992). Living in the country entails greater travel time to in-town tracks than living in the town where the track is situated. Wealth is associated with longer education, especially at school. Reading ability is known to be associated with school age and this is demonstrated by the research data where there are significant coefficients between age of leaving school and the measures of literacy.

The Thesis Model suggests that people who live more than 20 minutes from the track are less likely to become coaches than those who live within 20 minutes, i.e. those who live in the country are less likely to be coaches than those who live in the towns.

10.11.5. REJECTED VARIABLES - ATTITUDE TO VOLUNTEERING

Whilst this variable significantly contributes to the Thesis Model if added, it causes Tolerance Statistics to fall below 0.20. The largest correlation with the variable associated with the statement “Agree others should volunteer” is the variable related to the statement “Disagree coaches should be paid”. The latter is in the Thesis Model. This association is consistent with the traditional view that volunteering should be done without regard for personal gain.
If the variable “Attitude to others volunteering” is substituted in the model for the variable “Attitude to paying coaches”, there is a deterioration in the resultant model, probably because the latter attitudinal variable expresses more than just an attitude towards volunteering.

10.11.6. REJECTED VARIABLES - AGE

Both the variables related to age (“Age at Decision” and “Date of Birth” or “Generation”) can not be added to the model as they are correlated with several other variables in the Model and cause the Tolerance statistics to fall below 0.20. It is particularly associated with the number of cars available. This might be expected as car ownership increases with age as income and size of family increase.

The two age related variables are also correlated with two variables not in the model: “Household involved in sports other than athletics” and “% Post school years competing in sport”.

The older the respondent, the more the household is involved in other sports. This is probably because older people tend to be associated with bigger households and bigger the household, the greater the likely range of sports.

The older the respondent, the less post school competition there has been. This probably reflects the growing availability of and access to local sports facilities, particularly for women. The latter statement is born out by the fact the correlation coefficients of generation and for female respondents are 2 to 4 times higher than those for male respondents.
10.11.7. REJECTED VARIABLES - GENDER

An interaction effect with the variable “Gender” and the variable “Irregular work hours” is incorporated in the Thesis Model as the variable “Regular work hours and Gender”. A limited amount of correlation also exists between the variable “Gender” and other variables in the model, but not enough to cause significant collinearity, as measured by Tolerance statistics. But, if the variable “Gender” is added to the Thesis Model, the resultant collinearity becomes very significant and breaks the assumptions of logistic regression.

10.12. Conclusions from the Thesis Model

A logistic regression model with significant fit can be built from the experimental data. It indicates which explanatory variables are significantly associated with the decision to become a coach (the response variable) and the relative order of their importance. Some of these explanatory variables confirm the original hypotheses put forward. Others indicate that they are surrogates for or a result of interactions of some, unknown, more basic variables. The degree of dispersion also indicates that there may be one or more significant Explanatory variables that are absent.

However, many of the explanatory variables are not associated in a linear fashion with the response variable and require transformation. For most variables, the amount of data or its discrete nature only allowed a transformation to a categorical version, although this proved to be quite satisfactory for indicating the way in which the response variable changed with the explanatory variable.

The logistic regression model gives some qualitative and statistical credence to the conceptual Sports model put forward in Section 5.3. The Thesis Model shows that the incentive/disincentive effects of the variables are additive and that a fairly robust threshold exists between the decision being to become or not become a coach.
As discussed in Section 10.11. and elsewhere in this Section, the variable Gender is significantly associated with a number of other variables. Whilst some of the effect of this variable has been incorporated in the Thesis Model by use of the interaction variable, Genreg (see Section 10.5.), the Thesis Model could be extended to two models, one for each of the genders. This is discussed in Chapter 11.
11. GENDER MODELS

11.1. Introduction

The Thesis Model showed that the explanatory variable ‘Gender’ was significantly correlated to several other potential explanatory variables and that the effect of ‘Gender’ was probably not additive. This may indicate that the set of significant explanatory variables associated with male respondents may be different from the set associated with female respondents. In other words, the variables significantly associated with the decision by men to become a coach may be different to the variables associated by women.

The construction of two Gender Models was limited by the number of male and female respondents and the relative low number of female coaches. These restrictions limited the methods available to build models. The low numbers of cases in each gender caused mathematical problems in building models using either forwards or backwards entry to isolate significant variables from the potential variables. The numbers of cases available caused the standard errors of the calculated coefficients to be very high, particularly for the Female Model. As a consequence, the conclusions are restricted to general observations on the ability to use logistical regression to build acceptable models and on the variables in the model.

Initially, either all male or all female cases were selected from the total data set. The weight was set initial for those associated with all 112 cases. At any subsequent stage in building the Male Model, if any of the variables associated with literacy came into the Model, the weight was altered to that associated with the reduced 110 cases. This weight alteration was not needed for the building of the Female Model as all cases with data missing for the literacy variables were associated with male cases, not female.
The method used to build both models was to add a single explanatory variable and to measure the significance of the addition by calculating the value of chi-squared between the model with and without the variable and the significance of this. If the significance was greater than 0.010, the variable was rejected at that stage. If not, the Tolerances of all the variables in the model were calculated. If all Tolerances were greater than 0.20, the variable was designated as a possible entry. This procedure was repeated for every potential variable. When all variables had been through this procedure, the variable contributing the greatest significant log likelihood was added to the model. The building then commenced again and was repeated until no further significant variables could be added.

The model was then examined for the contribution of each variable now in the model. This was done by removing a variable and measuring the significance of its removal. The variable was replaced and the next variable inspected. The procedure continued until all variables in the model had been inspected. Those with significances greater than 0.20 were removed.

The re-building and removal processes were repeated until all variables in the model were significant. The resultant model was designated as the Male or Female Model.

The following two Sections, 11.2. and 11.3., give details of the Male and the Female Models respectively. Section 11.4. comments on the results and compares the two Gender Models.

11.2. The Male Model

The explanatory variable “Word Length” proved to be significant in this model. But the Box-Tidwell test proved to be significant indicating a non-normal distribution. To correct this, the variable was transformed using the square root. The variable “Square Root of Word length” was used in the Male Model.
The log-likelihood (-2LL) of the Initial Model is 776.616 and that of the Male Model is 73.977. The latter has 14 degrees of freedom and is significant at less than 0.001. The Nagelkerke R² coefficient has a value of 0.953, indicating that about 95% of the variance has been accounted for by the variables.

All 521 cases of “Not Coach” were correctly assigned and 82.3% of 39 cases “Coach” were correctly assigned. This assignment rate has a φ value of 0.838 that is significant at the probability level of 0.010. The C statistic has a value of 0.935, indicating that the assignment rate is better than chance.

The S statistic has a value of 5.264, indicating that there is considerable over dispersion in the model. Part of the reason for this may be the limited number of cases used to construct the model. Consequentially the covariance matrix used in the calculations in logistic regression may be sparsely filled as a result of limited examples of specific combinations.

All Tolerance statistics are above 0.200, indicating that no serious collinearity exists.

Of the 62 cases used to build the model, 6 have dBeta values greater than 1.00. Of these, only 3 have value over 1.98, the largest being 1871.11. This indicates that relative few cases exert strong influences on the values of the coefficients.

Diagram 11.2.1. shows the pattern of the calculated probabilities with the actual cases of “Coach” being indicated. The cut-off value of 0.5 is shown. This diagram indicates a satisfactory classification pattern with the cut off value not being unduly critical to the classifications.
Diagram 11.2.1. shows the distribution of the deviance values in relation to the expected normal distribution pattern, the theoretical distribution of deviances. It will be seen that the distortions from the Normal distribution mainly occur in the extreme cases of values approaching 1.0. These are the exceptional coaches having dBeta values above 2.00. The deviances are considered to be acceptable for a social model from a limited data set with some over dispersion.
The coefficients, their standard errors, the standardised coefficients, the Tolerance statistics and the -2LL contributed in the Male Model are shown in Table 11.2.1. It can be seen that the standard errors of the coefficients are high, relative to the 14 variables in the model, due to the low amount of data used in the model (from 62 cases).
Table 11.2.1.
DETAILS OF THE MALE MODEL

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Minutes to track (B)</td>
<td>8.888</td>
<td>3.814</td>
<td>0.466</td>
<td>0.594</td>
<td>47.539</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>Cars available (C) (1)</td>
<td>26.401</td>
<td>5.774</td>
<td>1.473</td>
<td>0.325</td>
<td>108.662</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>15.719</td>
<td>4.415</td>
<td>0.877</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time home (C) (1)</td>
<td>-6.552</td>
<td>1.634</td>
<td>-0.326</td>
<td>0.252</td>
<td>36.460</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>-5.681</td>
<td>1.536</td>
<td>-0.317</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys NOT in athletics (C) (1)</td>
<td>10.000</td>
<td>4.093</td>
<td>0.346</td>
<td>0.678</td>
<td>21.297</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>9.676</td>
<td>4.054</td>
<td>0.238</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys IN athletics (C) (1)</td>
<td>-8.353</td>
<td>4.266</td>
<td>-0.317</td>
<td>0.752</td>
<td>34.864</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>-2.217</td>
<td>4.002</td>
<td>-0.067</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls IN athletics (C) (1)</td>
<td>-11.404</td>
<td>2.752</td>
<td>-0.524</td>
<td>0.469</td>
<td>90.616</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>-2.894</td>
<td>1.862</td>
<td>-0.119</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SqRt Word length</td>
<td>-24.451</td>
<td>6.357</td>
<td>-0.004</td>
<td>0.203</td>
<td>37.372</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>Participation rate (B)</td>
<td>7.751</td>
<td>1.729</td>
<td>0.381</td>
<td>0.580</td>
<td>55.077</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>Work in Sport % (B)</td>
<td>12.686</td>
<td>2.578</td>
<td>0.478</td>
<td>0.651</td>
<td>50.594</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>NONE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTES to TABLE 11.2.1.

(B) = Binary variable

(C) = Categorical variable

If neither = Continuous variable

(1), (2), (3) refer to the categories of the variable

The coefficient of a variable is with respect to the reference category of that variable. The reference category is the last category. No coefficient is given for the reference category as this is allocated values of '0' in the coding.

A negative sign to a coefficient indicates that the variable decreases the probability of being a coach, in relation to the reference category.
The variables in Table 11.2.1. are now described.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TIME HOME</strong></td>
<td>20 minutes or less from home to track in relation to more than 20 minutes</td>
</tr>
<tr>
<td><strong>CARS</strong></td>
<td>(1) No or 1 car available in relation to 3 or more cars available &lt;br&gt; (2) 2 cars available in relation to 3 or more cars available</td>
</tr>
<tr>
<td><strong>HOME</strong></td>
<td>(1) Home by 4 p.m. in relation to home after 6 p.m. &lt;br&gt; (2) Home after 4 p.m. and by 6 p.m. in relation to after 6 p.m.</td>
</tr>
<tr>
<td><strong>BOYS NOT</strong></td>
<td>(1) No boys in household in athletic clubs in relation to 2 or more boys not in athletic clubs &lt;br&gt; (2) 1 boy in household in athletic clubs in relation to 2 or more boys in athletic clubs</td>
</tr>
<tr>
<td><strong>BOYS IN</strong></td>
<td>(1) No boys in household in athletic clubs in relation to 2 or more boys in athletic clubs &lt;br&gt; (2) 1 boy in household in athletic clubs in relation to 2 or more boys in athletic clubs</td>
</tr>
<tr>
<td><strong>GIRLS NOT</strong></td>
<td>(1) No girls in household in athletic clubs in relation to 2 or more girls not in athletic clubs &lt;br&gt; (2) 1 girl in household in athletic clubs in relation to 2 or more girls in athletic clubs</td>
</tr>
<tr>
<td><strong>GIRLS IN</strong></td>
<td>(1) No girls in household not in athletic clubs in relation to 2 or more girls in athletic clubs &lt;br&gt; (2) 1 girl in household in athletic clubs in relation to 2 or more girls in athletic clubs</td>
</tr>
<tr>
<td><strong>SQ RT WORD</strong></td>
<td>The square root of the maximum word length of leisure literature read by the respondent.</td>
</tr>
<tr>
<td><strong>PARTICIP.</strong></td>
<td>Involved in sport at least once a month in relation to involved less than once a month</td>
</tr>
<tr>
<td><strong>WORK</strong></td>
<td>More than 50% of work hours involves sport.</td>
</tr>
</tbody>
</table>
From the standardised coefficients in Table 11.2.1. it can be seen that one variable, “Cars available”, exerts a large effect, whilst another variable, “Square Root Word length”, contributes a small effect. These effects can be seen in Diagram 11.2.3.

In terms of fit, log-likelihood (-2LL), two variables contribute a large amount: “Cars available” and “Number of Girls IN athletics in the household”. The variable contributing the least is “Number of Boys NOT in athletics in the household”. The relative contributions can be seen in Diagram 11.2.4.
The same analysis follows in Section 11.3. to allow comparisons to be made in Section 11.4.

11.3. The Female Model

The explanatory variable “Number of Adults in Household” was added as, during the Model building, it was noticed that all other variables related to adults had some significance, but all were above the acceptable value of 0.010. The Box-Tidwell test proved to be not significant indicating a normal distribution.

The log-likelihood (-2LL) of the Initial Model is 681.505 and that of the Female Model is 25.598. The latter has 10 degrees of freedom and is significant at less than 0.001. The Nagelkerke $R^2$ coefficient has a value of 0.982, indicating that about 98% of the variance has been accounted for by the variables.
All 480 cases of “Not Coach” were correctly assigned and 74.5% of 16 cases “Coach” were correctly assigned. This assignment rate has a $\phi$ value of 0.828 that is significant at the probability level of 0.010. The C statistic has a value of 0.990, indicating that the assignment rate is better than chance.

The S statistic has a value of 2.560, indicating that there is a large amount of over dispersion in the Model. This is probably due to the low, 48, number of cases used and the very low number, 11, of coaches within these cases. Consequently, the covariance matrix used to calculate logistic regression models is too sparse with many empty entries for pairs.

All Tolerance statistics are above 0.200, indicating that no serious collinearity exists.

Of the 50 cases used to build the model, 3 have dBeta values greater than 1.00. The largest value is 12.39, which two cases have. This indicates that very few cases exert strong influences on the values of the coefficients.

Diagram 11.3.1. shows the pattern of the calculated probabilities with the actual cases of “Coach” being indicated. The cut-off value of 0.5 is shown. This diagram indicates a satisfactory classification pattern with the cut off value not being unduly critical to the classifications.

GENDER MODELS
Diagram 11.3.1. shows the distribution of the deviance values in relation to the expected normal distribution pattern, the theoretical distribution of deviances. It will be seen that the distortions from the Normal distribution mainly occur in the extreme cases of values approaching 1.0. These are the two coaches having dBeta values of 12.93. The deviances are considered to be acceptable for a social model from a limited data set with over dispersion.
The coefficients, their standard errors, the standardised coefficients, the Tolerance statistics and the -2LL contributed in the Male Model are shown in Table 11.3.1. It can be seen that the standard errors of the coefficients are extremely high, due to the low number of cases. This indicates that the values of the coefficients are highly unreliable. But the rank order of the variables, based upon their standardised coefficients, does give some indication of the relative effects of the variables.
Table 11.3.1.
DETAILS OF THE FEMALE MODEL

\[ \begin{array}{|c|c|c|c|c|c|c|c|} \hline \text{Variable} & \text{Coeff.} & \text{St. Error} & \text{Stand. Coeff.} & \text{Toler.} & \text{-2LL lost} & \text{Deg. of Free.} & \text{Sign.} \\ \hline \text{Cars available} & 63.588 & 306.32 & 0.368 & 0.315 & 59.628 & 2 & 0.000 \\ (C) (1) & 12.756 & 208.35 & 0.067 & & & & \\ (2) & 43.616 & 237.42 & 0.242 & 0.627 & 41.344 & 1 & 0.000 \\ \text{Irregular hours} & -16.291 & 306.52 & -0.089 & 0.573 & 27.672 & 2 & 0.000 \\ (C) (1) & -24.690 & 303.91 & -0.131 & & & & \\ (2) & -16.291 & 306.52 & -0.089 & 0.573 & 27.672 & 2 & 0.000 \\ \text{Girls IN athletics} & -40.246 & 189.48 & -0.233 & 0.411 & 28.610 & 1 & 0.000 \\ (C) (1) & -61.242 & 414.16 & -0.225 & 0.604 & 32.595 & 1 & 0.000 \\ \text{Adults} & -184.329 & 744.32 & -1.066 & 0.283 & 104.893 & 2 & 0.000 \\ \text{Pay Coaches?} & -38.360 & 182.36 & -0.191 & & & & \\ (C) (1) & & & & & & & \\ \text{Constant} & & & & & & & \\ \text{Notes to Table 11.3.1.} \\
(B) = Binary variable \\
(C) = Categorical variable \\
If neither = Continuous variable \\

(1), (2), (3) refer to the categories of the variable

The coefficient of a variable is with respect to the reference category of that variable. The reference category is the last category. No coefficient is given for the reference category as this is allocated values of ‘0’ in the coding.

A negative sign to a coefficient indicates that the variable decreases the probability of being a coach, in relation to the reference category.

The variables in Table 11.3.1. are now described.

CARS (1) No or 1 car available in relation to 3 or more cars available

GENDER MODELS
(2) 2 cars available in relation to 3 or more cars available.

IRREGULAR  (1) Work hours are regular in relation to being irregular.

GIRLS IN  (1) No girls in household not in athletic clubs in relation to 2 or more girls in athletic clubs

(2) 1 girl in household in athletic clubs in relation to 2 or more girls in athletic clubs

ADULTS  Number of adults, other than the respondent, in the household.

VET  (1) Not interested in veteran athletics in relation to being interested or having competed.

ROAD  (1) Main interest is track & field or cross country, in relation to main interest being Road Running.

PAY  (1) Agree with statement that “Coaches at local sports clubs, generally, should be paid”, in relation to Disagreeing.

(2) Neutral to statement that “Coaches at local sports clubs, generally, should be paid”, in relation to Disagreeing.

From the Standardised Coefficients in Table 11.3.1. it can be seen that one variable, “Attitude to paying coaches”, exerts a large effect. This effects can be seen in Diagram 1232.3.
Diagram 11.3.3.

STANDARDISED COEFFICIENTS - FEMALE MODEL

In terms of fit, log-likelihood (-2LL), one variable, “Agreement to paying coaches”, contributes a very large amount of the fit. The relative contributions to fit of all the variables can be seen in Diagram 11.3.4.
In the next Section, 11.4., the explanatory variables in the Male and Female Models are compared.

### 11.4. Gender Model Comparisons

Diagram 11.4.1. compares the percentage of fit contributed by individual explanatory variables in the Male and the Female Model. The percentage fit was calculated as:

\[
\text{% Fit of Variable} = \frac{-2LL \text{ contributed by variable}}{\text{Sum of -2LL of all variables}} \quad \text{(Equation 11.3.1.)}
\]
There are major differences between the set of variables in each model with few common variables. Only two variables, "Number of Girls IN athletics" and "Number of cars available" are common to both models.

Table 11.4.1. makes a comparison of the signs of the coefficients in the two Gender Models. Section 10.10. showed that, where a logistic regression model has no constant, the signs of the variables indicate whether the influence of that variable is an incentive (positive sign) or a disincentive (negative sign). In the case of categorical variables, the Table shows whether or not coefficient values for the classes move towards zero (declining) or away from zero (increasing). Discrepancies with the hypothesised or expected states are emboldened. These discrepancies are discussed after the Table.
Table 11.4.1.
COMPARISON OF INCENTIVE/DISINCENTIVE EXPECTATIONS WITH THESIS MODEL

<table>
<thead>
<tr>
<th>Variable Group</th>
<th>Incentive or Expected</th>
<th>Male Model</th>
<th>Female Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>No or 1 Dependant NOT in athletics</td>
<td>Hypothesised or Expected</td>
<td>Incentive Positive Increasing</td>
<td>Incentive Positive Declining</td>
</tr>
<tr>
<td>(compared to 2 or more)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No or 1 Dependant IN athletics</td>
<td>Incentive Negative Declining</td>
<td>Disincentive Negative Declining</td>
<td>Disincentive Negative Declining</td>
</tr>
<tr>
<td>(compared to 2 or more)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adults in Household</td>
<td>Incentive Positive Increasing</td>
<td>Incentive Positive Increasing</td>
<td>Incentive Positive Increasing</td>
</tr>
<tr>
<td>(increasing number from None)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 &amp; 1 or 2 cars available</td>
<td>Incentive Negative Declining</td>
<td>Disincentive Negative Declining</td>
<td>Disincentive Negative Declining</td>
</tr>
<tr>
<td>(compared to more than 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular hours</td>
<td>Incentive Positive</td>
<td>Incentive Positive</td>
<td>Incentive Positive</td>
</tr>
<tr>
<td>(compared to irregular)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home by 6 p.m. &amp; by 7 p.m.</td>
<td>Incentive Positive</td>
<td>Disincentive Negative Declining</td>
<td>Disincentive Negative Declining</td>
</tr>
<tr>
<td>(compared to after 8 p.m.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not interested in Veteran Athletics</td>
<td>Incentive Positive</td>
<td>Disincentive Negative</td>
<td>Disincentive Negative</td>
</tr>
<tr>
<td>(compared to interested)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main interest NOT Road Running</td>
<td>Incentive Positive</td>
<td>Disincentive Negative</td>
<td>Disincentive Negative</td>
</tr>
<tr>
<td>(compared to Main Interest)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participate 1/month or more</td>
<td>Incentive Positive</td>
<td>Incentive Positive</td>
<td>Incentive Positive</td>
</tr>
<tr>
<td>(compared to less)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word complexity</td>
<td>Incentive Positive</td>
<td>Disincentive Negative</td>
<td>Disincentive Negative</td>
</tr>
<tr>
<td>(Increasing)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 50% sport in work</td>
<td>Incentive Negative</td>
<td>Disincentive Positive</td>
<td>Disincentive Negative</td>
</tr>
<tr>
<td>(Compared to at least 50%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strongly Agree or Neutral to paying</td>
<td>Disincentive Negative Increasing</td>
<td>Disincentive Negative Increasing</td>
<td>Disincentive Negative Increasing</td>
</tr>
<tr>
<td>coaches</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The discrepancies related to “Number of Cars available” and “Hour Home” were discussed previously in Section 10.9.2. in relation to the Thesis Model.

In the case of the Male Model, the apparent decline in the class coefficients of the categorical variable associated with number of boys NOT in athletics may not be significant as the standard errors of the coefficients are several times the difference in the values of the coefficients for the classes. Only a much larger sample could show whether or not there is a significant decline. If the decline was significant, a logical explanation may be that when a male guardian has no sons not involved in athletics,
everything else being equal, he may feel free to take up athletics coaching. When
there are one or more sons not involved in athletics, the male guardian will have
incentives to involve himself with them at the expense of any incentives to become an
athletics coach.

In the case of the Female Model, the standard errors of the coefficients are such that
there are no statistical differences between the calculated values of the coefficients
and zero. Thus, the apparent differences between the hypotheses or explanations may
not be significant and may be due purely to the sample size.

11.5. Conclusions from Gender Models

Four conclusions can be drawn from constructing the Gender Models:

i. The sets of explanatory variables for the two genders differ, both in the
variables within each set and in terms of the importance of individual
variables. In particular, variables associated with Attitudes and Interests
appear to be more dominant in the Female Model than in the Male
Model. But the absence of a variable in one gender model does not
indicate that that variable plays no part in the model, only that the effect
found from the data used was either too small to be statistically
significant or was masked by covariance with other variables. One
example of this may be shown by the fact that interest in veteran
athletics is a significant variable in the female model, but not in the male
model. Interest in veteran athletics may be more significant with women
because the age for becoming a veteran is younger for women than for
men. Consequently becoming a veteran is more accessible to women
than to men.

ii. Variables associated with Cars, Time home and the number of children,
especially girls, involved in athletics appear to be important in both
Models. The latter is intuitively acceptable and indicates the influence of family interests in decision making of the parents. The strong negative influence of the number of cars and the strong positive influence of the time home from work is contra-intuitive. The number of cars may be a surrogate measure of socio-economic status.

iii. In neither Gender Model does there appear to be a significant effect of the Club. This is not due to correlation between Gender and Club. This can be demonstrated by the following. As both these variables are binary, the association between Gender and Club can be measured by calculating Cramer’s C (Black, T R, 1999). The value is 0.046, which is not significant at a confidence level of 0.010. This illustrates the value of the Gender Models rather than the non gender Thesis Model. In the latter, Club is a significant explanatory variable (see Table 10.5.2.).

iv. Gender Models built with the limited number of cases available in this research can only be used as indicators of significant explanatory variables and of their rank order of importance. The calculated values of the coefficients are unreliable due to the high standard errors. The number of cases required to obtain reliable values of the coefficients would probably need to be of the order of 150-200 for either Gender Model.

In addition to these conclusions, there is some indication that there are two distinct populations entering the Coach Education Programme:

- Those who whose prime intention is to coach regularly at a club when they have qualified - Leisure motivated coaches.

- Those whose prime incentive is to obtain a qualification to enhance their career and who no immediate intention of coaching regularly at a club - Career motivated coaches. These do not, strictly, meet the definition of ‘Volunteer’ used in this

GENDER MODELS
research and shown in Section 2.2. However, they are becoming qualified coaches without expectation of direct payment for being a coach.

In the same way that ‘Gender’ is not an additive Explanatory variable and that the two genders have differing sets of significant Explanatory variables, it is possible that Career motivated Coaches may have a different set of associated Explanatory variables to Leisure motivated Coaches. The indications are that if there is such a difference, it may be separate and in addition to the gender division, i.e. there may be 4 populations: Male Leisure, Male Career, Female Leisure and Female Career. The research sample was too small to have sufficiently large numbers in each of these four categories to allow further modelling.
12. QUALITATIVE EVIDENCE FROM INTERVIEWS

12.1. Introduction

The second half of each interview explored with each respondent their association with athletics, their views on coaching and the circumstances under which they might coach themselves. Whilst the main reason for collecting this information was to inform the clubs participating in the research, it was hoped that the information would also supplement and, perhaps, corroborate, the quantitative model. It was also hoped that some indications might be obtained on the ‘Key’ events which triggered decisions to enter the Coach Education programme, together with the timing of such ‘Key’ events.

Each recorded interview was analysed to extract all statements made by the respondent related to how they perceived coaches in athletics, what they expected of coaches and why they had become coaches or why they were not coaches. The statements were then grouped into broad areas and analysed by the following factors:

- gender,
- age,
- marital status,
- form of association with athletics (member or parent),
- reasons for association (personal participation, child’s participation),
- interest in veteran athletics,
- interest in road running,
- perception of coaches (ex athlete or not)

The analysis could not be and was not rigorous, in a statistical sense, as the second part of each interview was conducted in a very loosely structured manner as a conversation. The ground covered by each interview varied, as did the amount of time and depth spent on each topic. The results were intended purely to supplement and amplify the main quantitative modelling.
12.2. **Evidence from Non-Coaches**

Non-Coaches are defined as members and parents who are not and have not been qualified athletic coaches, i.e. have never entered the Coach Education Programme and do not or have not held an athletics coaching qualification.

76 interviews with Non-Coaches were recorded and analysed from 38 male and 38 female respondents with 38 from Club A and 38 from Club B. Of these 76 (14%) said that they might become athletic coaches sometime in the future and 28% said that they would never consider becoming a coach. The majority, 58%, did not reject the idea of becoming a coach, but had no strongly held views or ideas on the subject. Those who expressed the greatest interest in becoming coaches in the future lay in the age range of 35 to 44 years old. This age range corresponds to the ages at which the majority of existing, active coaches first registered. Men were more inclined to consider coaching than women.

5 predominate reasons were given for why a respondent was not or would not become a coach, together with a 5th category of sundry reasons.

i. 34% of respondents stated that they already had one or more leisure interests outside athletics that took up their spare time. Often these other activities involved one or more other members of their family and had often been established before they had become associated with athletics.

ii. 22% of respondents said that they had no time spare as a result of their work (especially those in professions such as teachers and doctors) or running their own business or frequently having to travel abroad or being a single parent or from family pressures of children or elderly or sick relatives.

QUALITATIVE EVIDENCE
iii. 22% of respondents actively competed or trained in athletics and stated that their priority was to continue active participation as long as they were able to do so. This included being a veteran athlete. They would only consider becoming a coach when injury or age made further physical participation either impossible or unattractive. Most saw no possibility of giving up competition in the foreseeable future.

iv. 9% of respondents felt that they could not commit the regular frequency and amount of time that they perceived athletic clubs expected from coaches. Sometimes this was because they never knew when they would finish work and had minimal control over their hours of work. This group might represent a future potential for recruitment if the regularity of commitment could be relaxed. Some respondents thought that they might be interested if coaching was organised in a pool system. In such a system, absence would be covered by another coach in the pool.

v. 12% of respondents gave other reasons, the principle one being that they were not interested in sport of any kind. This reason was given by several of the women. Another set of reasons given by several respondents was that they did not want to take the risks that they perceived to be in coaching. Risks mentioned included being wrongly accused of sexual abuse (by men), selecting the wrong training and being subjected to bad behaviour (especially from young boys).

Consideration of becoming a coach was also coloured by perceptions. 67% per cent of all parents said that they thought that all coaches were ex-athletes. As most parents had not been athletes themselves, their perception was that they did not have

QUALITATIVE EVIDENCE
the right background or knowledge to be accepted as an athletics coach. Overall, 47% of respondents, including members, assumed that athletics coaches had competed in athletics themselves at some level. This perception was held more strongly by women than by men. Some practising athletes had more confidence in somebody who had personal experience of an event rather than somebody who had the technical knowledge but not the specific physically experience. Where a respondent was a member, but not an athlete, this view became a deterrent to them taking up coaching, but not to becoming an Official or an Administrator or Team Manager.

There was also a tendency for mothers to be more interested in coaching their sons, but not so frequently their daughters. To a lesser extent, fathers tended to be more interested in coaching their daughters, rather than their sons. It was not the purpose of this research to investigate this tendency and the interviews shed no light on the matter.

What came out very strongly in the interviews was the universal lack of knowledge about the coaching system, the education programme and the qualifications. This applied almost as much to existing members of athletics clubs as to the parents who were not members. Only in a few cases did respondents say that they had been approached to become coaches. Yet almost all of them considered coaching to be important or very important and almost all considered that coaches should be formally qualified. However, coaching was seen to be mainly concerned with teaching the technical elements of athletics to young children or for athletes who were exceptionally talented. In particular, veteran athletes and road runners tended to think that their events did not need formal coaching. In the case of veterans, they tended to think that their experience as younger athletes was sufficient guidance. In the case of road runners, they tended to consider that their colleagues with whom they ran could give sufficient advice.
Very few respondents had known in advance if there was coaching at the athletics club before they or their children joined. Even where they had some knowledge in advance, it was second hand from friends or from teachers. Very often parents assumed that as the track facilities were owned by a local authority, there must be coaching and that this must be by qualified coaches. Generally parents said that they had been down and observed the first few sessions and had been satisfied that the coaches were satisfactory. Virtually no respondent had made formal enquiries on the qualification of coaches in general or the qualifications of a specific coach. This was born out by the coaches who all said that they had hardly ever been asked about their qualifications. Only one respondent mentioned the subject of Child Protection. This was a parent who had become a coach as a result of fears of an existing coach.

Most respondents were very appreciative of the coaching that was available, particularly that individual coaches could be relied upon to always be at every session. This was reflected in their perception of the time commitment required of a coach. They considered that interpersonal communication skills were just as important in a coach as knowledge of athletics.

Many respondents saw the voluntary or unpaid nature of coaching to be of positive benefit for a number of reasons. It was felt that payment changed the motivation of coaches away from helping all to concentrating upon those with perceived talent. There was concern that payment would increase the tendency for some coaches to push young children too hard and to concentrate upon gaining short term competitive successes at the expense of the longer term development of each individual to his or her potential.

There was a significant minority of respondents who felt that the general organisation of clubs, especially the coaching, was unsatisfactory and haphazard and that one or more professional coaches in each club would significantly improve the organisation and quality of coaching. Some of the younger respondents felt that the shortage of coaches was partly due to it being unpaid. Several young parents remarked to the
effect that “nowadays you only get what you pay for”. There was almost uniform
concern that any move to paying coaches would disadvantage many young athletes
and might increase club fees so much that young people starting on low wages or
studying could not afford athletics.

When the benefits of paying coaches were discussed in some depth, there was almost
universal agreement that it would be impossible to pay some coaches and not others,
even on the basis of qualifications or additional duties such as overall organisation or
mentoring. The only exception might be occasional visits of outside experts by
invitation of the coaches at the club. There was a fear that the appointment of paid
people in athletics as development officers would not only further deter young people
from taking up voluntary coaching (and other posts in clubs), it could also demoralise
existing volunteers (including coaches) if they felt their freedoms of action being
reduced by bureaucracy. Some respondents felt that this was already happening.

The biggest perceived weakness of the existing coaching systems seemed to be the
lack of coaches, in particular the effects of coaching groups larger than 5 or 6 people.
As will be seen in the next Section (12.3.) (Evidence from coaches), this is one of the
main reasons why parents, in particular, start coaching. More experienced athletes
and their parents were critical of the lack of coaches with more advanced knowledge
of specific events and the lack of facilities to locate and contact suitable coaches
within the County. The role of professional coaches was seen as being important if it
could improve the quality of athletics teaching at schools and the rate at which
talented children could be spotted at school and directed to the local club.

There was a dichotomy of desire from parents of both clubs. One set of parents was
full of praise of both clubs for the way that they took in young children and first
insisted that they tried all the events without too much pressure, especially of
competition. Many parents wanted this to continue at least until children were 14 to
15 years old. These parents set great store on the social aspects of sport. Another set
of parents wanted this introductory phase to be shorter and for their children to

QUALITATIVE EVIDENCE
quickly receive coaching in a specific range of events in which their child expressed interest. These parents set great store on the social value of learning to be competitive and of personal achievement. Both types of parent reported that other of their children had left the athletics club either because they felt too much pressure to do a single event (possibly separating them from their friends in other event groups) or because they wanted to specialise and could not do so.

In summary, coaching was highly associated by respondents with the participation of children rather than the participation of adults. Personal participation in athletics was perceived as a significant prerequisite to being a coach. Almost none of the parents had been athletes themselves and most of them were involved with one or more other sports or social activities. Most of their children were involved with other sports.

For respondents training and competing in athletics, coaching was seen only as a means of remaining involved after they ceased competing. With the advent of veteran athletics, the only reasons for ceasing competition were seen as long term injury or declining health. Coaching and competing were not seen as being compatible. These views were especially held by younger respondents. Only payment for coaching might change their intentions. Recognised qualifications in coaching were, primarily, seen as benefits for career and CV purposes. They admitted that they had little interest in regularly coaching until they had ceased competing.

The biggest deterrents to becoming a coach were the pressures from work and, in the case of women, looking after a family. Only four women stated that they did not have a paid job or were not studying. Most women reported that they had a full time paid job. Most respondents said that it was the intensity and uncertainties of their work that deterred them from volunteering, not their actual work hours. In many cases, respondents either had work that had to be done at home or they could not guarantee at what time they could leave work. Consequently, many of them felt that they could not give the regular commitment that they perceived as being necessary to coach.

**QUALITATIVE EVIDENCE**
12.3. Evidence from Coaches

The interviews with 34 coaches could be analysed for the reasons why they started coaching and for their views and perceptions of coaching. 68% were men and 32% women. 56% had only a Level 1 qualification. The vast majority of the remaining 44% had a Level 2 qualification.

26% of coaches were under 35 when they entered the Coach Education Programme. 35% of coaches were between the ages of 35 and 44 when they entered and a further 32% between the ages of 45 and 54.

32% of all coaches were not regularly coaching when interviewed. Of these, 50% were those under 35 years old.

The main reasons given for coaching could be put into 5 categories.

i. A third of respondents started coaching because they observed gaps or shortages in the coaching resources at the club, but not directly related to the events in which their own children participated. The majority of these had become associated with the club through their children. Some of these had started coaching to occupy their time at the track waiting for their children. This reason is returned to in Section 12.4., `Key’ events.

ii. A quarter of coaches stated that they entered coaching because their own children had joined a club and they wanted to help them. The majority felt that the existing coaching was not adequate, often because either the coaching groups were too big or because a specific event was not adequately covered or, in the case of boys, there was not a coach at all. As noted before, this group tended to...
be fathers with daughters as athletes or mothers with sons. This reason is returned to in Section 12.4., ‘Key’ events.

iii. 1 in 5 entered the coach education programme to gain a qualification for their own career development. These were evenly split between young people wishing to improve their CV’s for entrance to Further or Higher Education and professional teachers or professional health instructors. The majority of these respondents did not regularly coach at clubs after they had gained their Level 1 qualification. In the case of the younger respondents, their own competitive career took priority. In the case of professionals, they used their time at their clubs as a relaxation from instructing, especially teachers. The proportion of these mainly inactive (for clubs) coaches was nearly 50% with those under 35 years old and fell as the age group rose. There was an equal proportion of men and women in this category. This reason is returned to in Section 12.4., ‘Key’ events.

iv. 1 in 8 of respondent coaches stated that they had finished competing and wanted to put something back into sport. Some of them had been athletes and they saw coaching as a means of retaining social contact within the club. This reason showed signs of declining as the opportunities increased for continuing competition as a veteran or by road running. This was reflected in the responses of the non-coaches (Section 12.2.) in which athletes put higher priority on continuing competitive competition than on coaching.

v. The remaining respondents gave other reasons, such as wanting to remain involved after they had ceased competing and doing something whilst they waited for their children at the track.
There was a tendency with the Level 1 coaches to not have any intention to progress to Level 2. Three main reasons were given:

i. They were not prepared to give further time to courses as this would be at the expense of the athletes they were coaching and they felt that they either had sufficient knowledge or that they could gain further knowledge from observation and reading. Many of them felt that the training they had received at Level 1 was quite adequate for the situations they encountered and that their athletes were gaining success with their present training. The majority of these coaches appeared to be coaching mainly young athletes under the age of 15.

ii. They did not like formal instruction and examinations or assessments. In some cases, they had reluctantly taken their Level 1 qualification to be covered for insurance. In some cases, revalidation may cause them to either cease coaching or to continue coaching without a valid licence. The policy of each club towards the latter was not known.

iii. Most of those who had taken the Level 1 qualification for career enhancement felt that they would not gain from a more advanced qualification, largely because they had now been accepted into Further or Higher Education, the original objective of enhancing their CV.

All but 2 of the coaches said that they did not regret having become a coach and would do so again. However, there was a significant amount of concern and disillusionment expressed about current developments. These arose from three areas.

i. The growing tendency for parents to, apparently, take no interest in the activity of their child or children at the club. "We are just seen
as unpaid child minders. They drop them out of the car and shout for them to go home from the car. We never see them otherwise.” Coaches resented that some children came regularly to training during the week and then were not allowed to compete at the weekends as it was not convenient to the family to take and collect them before and after. Worse still, children agreed to attend a competition and then did not, without ever giving a reason or apology. Other children came irregularly, sometimes because of other family activities. The behaviour of children, especially young boys, was frequently referred to and unacceptable behaviour was associated with apparent lack of interest by the parents.

ii. Lack of communication from the sport. Many coaches regretted that the professional regional coaches appointed during the period of the BAF had not been replaced. The comments of the coaches showed the high respect in which the regional coaches were held. The ability to be able to meet them, work with them and to contact them over the telephone was considered to be invaluable and highly supportive. Regional coaching days, county coaching forums and being asked to attend courses were much appreciated. Development officers were not held in such high esteem and were deemed to be less accessible and not so knowledgeable. A frequent comment was “Once you have taken your qualification, nobody seems to be interested in you and you hear nothing from outside the Club”. Coaches wanted opportunities to meet coaches from other clubs. Coaches and their athletes needed some local (county or regional) focal point for advice and information.

iii. Most coaches expressed concern about what they perceived as a growing amount of external bureaucracy and interference in an activity which they that they had freely and voluntarily entered for
their own enjoyed and for recreation. This concern appears to have arisen from issues surrounding coaching qualifications (especially revalidation, upgrading and insurance); child protection (personal checks; instructions relating to touching children, taking them in cars, etc.); and health and safety (dealing with injuries, changes in equipment specifications). In the same vein, there was some disillusionment arising from the decline in financial support from Local Authorities and the perceived difficulty in applying to and getting funding from the National Lottery. There was also resentment expressed at the increasing number of paid administrative posts in athletics and sport employed by the NGB’s and local authorities. Not only was there some feeling that these people were beginning to tell clubs what to do, the posts were absorbing money that the coaches felt would be better spent in financially helping their club. The resentment was summed up by one respondent: “I come down to coach to escape the increasing bureaucracy I suffer every day at work. I want to do my own thing”.

In summary, the majority of respondents may have become coaches not because they were altruistic, but because they saw benefits to either themselves or to their immediate family. The often stated desire to give something back to sport may be as much a desire to remain associated with sport than altruism. This desire may develop the longer the respondent coaches. Coaching may well become an important means of escaping the stresses of work and life. This could be one explanation of the resentments associated with perceived increases in bureaucracy and the perceived threat to a volunteer from the increases in paid and professional staff in sport. The emphasis on communication and on dependency from a coaching group may be expressions of a need for recognition. This may be associated with the fact that many respondents started coaching because they were asked to do so. All these observations are potential areas for future objective research.
12.4. ‘Key’ Events

Whilst the research was not designed to identify the ‘Key’ event that triggered a
decision to become a coach, the interviews did give some indication of what these
might be.

A third of coach respondents had been asked by a member of the club to become a
coach. It is probable that other coach respondents had been asked, but did not recall
this at the time of interview. The most frequent person to ask a respondent to
become a coach was the initial coach of their child. Often this was because they either
wanted help with too large a group of athletes or they were ceasing to coach because
of business commitments or ill health.

Interestingly, a number of respondents who were not coaches mentioned that they had
been asked by officials in the club to become coaches. Many of these were teachers.
The quantitative model indicated that the net value significant variables associated
with becoming a coach had not reached the necessary threshold. In contrast, other
respondents who were not coaches remarked that they might have become coaches in
the past if they had been asked and the Coach Education programme had been
explained to them. When asked if they might become coaches in the future, the reply
was often that they felt it was too late either because they were too old or because
they had taken up other activities.

This illustrates the importance of being asked and the importance of the timing of
asking.

Another ‘key’ event mentioned by coaches was that their child had reported to them
that they were unhappy with the personality or the depth of knowledge of their
existing coach. Parents had sometimes found difficulty in either identifying another
coach, especially if the problem was event specific knowledge, or felt uncomfortable
about asking the club to change the coach. The parent might then decide to become
coach in order to solve the problem. This may also have been the situation with those coaches who stated that they started because of a perceived shortage of coaches.

An athlete’s dissatisfaction with a coach as a potential ‘key’ event was also illustrated by comments from respondents who were not coaches. In such cases, respondents often gave finding a suitable coach as a main weakness in the existing system. They sometimes commented on the adverse reactions which they received from the club when they considered getting coaching assistance from outside the club. Some parents reported that the inability to solve the coach problem had led to their child withdrawing from the sport.

This illustrates the close association that can exist between ‘key’ events and factors associated with changes in motivational levels. Quality of relationship with officials in the club, including coaches, may be a significant Explanatory variable. It does illustrate how important it is for clubs to monitor relationships and to facilitate solutions where relationships are failing.

The other ‘key’ event which was reported was related to the Life Cycle Event of entering or changing career. This could have two effects. One was to enter the Coach Education programme in order to gain a qualification for a CV. This only occurred when the career related to sport or leisure. The other reason led to the opposite result. The entry to a career or the need to change career were sometimes offered as reasons why a respondent had insufficient time to take up coaching.

Whilst taking a coaching qualification solely to enhance a CV may have no short term benefit to sport, it can be viewed as a long term investment. However, this investment will only be realised if the sport can keep in communication with the individual until the individual can become active. This raises questions as to how the current revalidation rules can be adapted to a situation where an individual may be temporarily prevented from taking an active coach role for a number of years.
12.5. Comparing the Qualitative Evidence with the Quantitative Models

Trying to understand a population from qualitative data and from quantitative data can be compared to looking into a room from two different windows. The different viewpoints and perspectives are bound to give different images with some aspects more dominant in one perspective than in the other and others aspects only visible from one perspective. Nevertheless, it would be expected that each perspective would confirm many items in the other perspective and that the overall picture would be similar.

This Section attempts to compare the perspective of coaching gained from the quantitative models with the perspective obtained from the qualitative data from the open ended part of each interview.

All the quantitative Models showed very strong, but negative, associations with coaching and the number of cars available. This seems to contradict the hypothesis that restricted access to personal transport will restrict the opportunity to volunteer. This was never mentioned as a problem in any interview. However, number of cars may be an indirect measure of one or both of two other factors:

- The number of adults in the household and, hence, the potential range of household interests and activities other than athletics. Such an association probably exists as the correlation coefficient between Number of Cars available and the Number of Adults in the Household is 0.360 (N=112). A negative relationship between being a coach and the household having many other interests was qualitatively indicated in the interviews by the large number of respondents that said that other activities prevented them from helping. Also some coaches referred to other interests that they had sacrificed in order to coach. In some cases, this was a reason why they had ceased coaching or were considering doing so.
The socio-economic status of the household with the wealthier categories having more cars. Whilst there is no known research, indications are that participation in athletics in the UK is less associated with the wealthier socio-economic sections of the population than with the less wealthy. If this is the case, those being active in the sport, including coaching are less likely to have several cars available to them. This would also account for the fact that many respondents disagreed with paying coaches on the basis that this would disadvantage households on low incomes. The socio-economic aspect of athletics may also be related to the slight negative association between word length and the probability of being a male coach.

In all the models, the greater the participation of children in athletics, the greater the probability of the respondent becoming a coach. This was particularly true where the children were girls. This may be because, overall, respondents had 50% more girls than boys in their households. The association between coaching and children was strongly borne out in the interviews, both in terms of why a respondent might coach and what the respondent felt was the most important role of a coach. Interviews also indicated that men were more likely to take up coaching if they had a daughter involved than if they had a boy involved. The models tended to reinforce this gender bias. The majority of coaches stated that they had become involved with athletics following one or more of their children joining and because they wanted to, directly or indirectly, help their children. Similarly, some respondents showed little interest in athletics and coaching because the majority of their children were not involved with athletics.

Analysis of the interviews indicated that women tended to be much more concerned than men about the social benefits of athletics to children, in terms of learning relationships. They were also more concerned than the men about coaches
pressurising children. If these concerns are allied to their perception that paying a coach may increase pressures on the children, then the strong positive association in the Female Model can be better understood between strongly disagreeing with paying coaches and the probability of being a coach. It could explain why some women coach as unpaid volunteers in an attempt to protect their children from such pressures, i.e. ensuring that paid coaches are not involved.

The Male Model had a negative association between how far the respondent was from the track and the probability of being a coach. The Female Model had a negative association between irregular work hours and being a coach. Both these factors, distance and irregularity, can be seen as pressures on free time. This was reflected in the qualitative data from the interviews were lack of time from business and family commitments were a frequent reason put forward for not becoming a coach.

In the case of the Male Model, but not in the Female Model, the amount of sport involved with a respondent’s work was strongly associated with becoming a coach. As has been observed, becoming a coach can be solely to gain a qualification for career purposes, but without any immediate intention of actively coaching. Overall about 1 in 4 of all respondents reported that they had some sport involved with their work or study. But, in the case of women, the majority had work involved with young children, mainly in teaching. Many of these respondents did not want their leisure to be concerned with children and, thus, did not want to become coaches. Very few male respondents were teachers and the sport in their work was more concerned with products from their work or services mainly to adults. Continuing sport into their leisure activity working with children is not a problem for the men as they do not have continuous contact with children as do women through teaching. Indeed their leisure participation in sport may well enhance their business contacts and knowledge.

In interview, women expressed much stronger views than men on the subject of veteran athletics and on road running. Both these are mirrored in the Female Model

QUALITATIVE EVIDENCE
where there are negative associations between an interest in either and the probability of being a coach. In the case of the Male Model, there is a strong positive association between the frequency of sports participation and the probability of being a coach. Unlike the women, this participation includes being a veteran athlete or a road runner. In the interviews, it was more male respondents than female respondents who expressed the view that a coach is likely to have been an athlete. It then follows that if a man participates himself, he is more likely to see himself as a coach and be accepted as such by athletes.

In summary, several of the dominant features of the qualitative data from the interviews can be logically associated with the variables that appear in the quantitative models, albeit that slightly different definitions of aspects have to be taken have to be taken. This corresponds to the window analogy given in the opening paragraph of this Section. There is sufficient correspondence between the quantitative and the qualitative results to suggest that each supports the overall viewpoint of the other.
Much quantitative research on why people decide to enter an educational programme has been based upon a premise that the variables used are continuous, normally distributed and not significantly correlated. Conclusions have often been arrived at by using statistical tests such as linear regression and Chi-square that have assumptions about one or more of the latter three characteristics. Much sports research has been based upon limited or generalised theoretical models, little has involved parents and the limited work on coaches has been on professional and not voluntary coaches. This research has attempted to improve all these situations. Although a limited field has been used for the research, decisions on coach education in two clubs and in one sport, the aim was to make and to justify advances in modelling which can be extended to other areas of education, reading and sports research.

It needs to be emphasised that this research was to aid identification of associations. Other types of research are required to show whether or not the variables identified are causal or not and the mechanisms of causality. The research may also aid prediction of the probability of sports and leisure related decisions.

This Chapter starts by examining the extent to which the research question set out in Section 1.1. has been answered.

13.1. Answering the Research Question

The research question originally posed was:

"DOES THE DECISION TO ENTER THE UK ATHLETICS COACH EDUCATION PROGRAMME CONFORM TO A LOGISTIC REGRESSION MODEL?"
A further question arises from this as to whether or not such a statistical model fits into any of the published general models of decisions in sport and recreation.

The Thesis Model shows that a statistically significant logistic regression model can be built from a set of personal factors and attitudinal variables suggested by existing published research on sport, coaching and volunteering. However, examination of the characteristics of this model indicates that the set of explanatory variables may vary between the genders. This was shown to be true by constructing a Male and a Female Model from the same data as the Thesis Model.

The fact that all three Models had significant amounts of over dispersion suggests that further explanatory variables exist. It was also shown that some of the variables suggested by existing research are complex variables that do not reveal the basic underlying factors making them up. One example is “Number of cars available” which is suggested in some published literature as a measure of the ability of a person to access sports, leisure or educational sites.

The research has been able to extend existing published decision models of leisure to the specific case of joining a sports coach educational programme (see Chapter 5). These models are based upon the concept of incentives and disincentives that raise or lower the probability for an individual of exceeding a threshold value. These incentives and disincentives work in an additive mode, disincentives having negative values. A logistic regression model assumes that the effects of all variables are additive. The models developed during the research show that, within a wide range, the default threshold value used in logistic regression is not critical. The research covered a range of individuals and a range of time. This suggests that the modelling is fairly robust with respect to both these factors.

The research also indicates that, for at least some variables, the effect of a variable may not be linear and continuous. Rather the effect reaches a critical value for that variable and then remains either constant or continues to rise only at a low rate. An
example of a variable reaching a constant may be the time the respondent perceives to be from the athletics track. This variable appears to be a disincentive. Once the distance exceeds a critical value, about 20 minutes, the resultant disincentive becomes very large and remains so for all greater times. An example of a variable that tapers out in effect is the number of girls in the household involved in athletics. Having one girl involved appears to be a major incentive. Having more than one girl involved has some additional incentive, but not as great as the first.

13.2. Limitations of the Research

The main limitation of this research is that the models are only based upon two similar Clubs. Whilst the Thesis and Gender Models indicate that the variable “Club” is not significant in its own right, it is very likely that there are geographic differences. For example, it is probable that the distances which people are prepared to travel for their sport may be related to the distances they have to travel for other non-work purposes, such as shopping and schools. These distances will be at a minimum in large cities and at a maximum in sparsely populated rural areas.

The limitations resulting from the population sampled also include ethnicity. The catchment areas of both the clubs used in the research do not contain significant numbers of the ethnic populations chiefly associated with athletics, especially the afro-caribbean population. Both catchment areas tend to have a greater frequency of the high socio-economic groups. Both catchment areas also have a high proportion of people who commute significant distances to work as a result of both areas being well serviced by fast trains to London. Similarly, because of their proximity to Heathrow and Gatwick airports, they have a higher than normal proportion of people who fly on business and are away during weekdays. The areas also have very low unemployment rates.

It is not clear how a variety of geographic areas could be handled by logistic regression. Traditionally, area would be introduced as another, probably categorical,
variable. However, two assumptions in logistic regression are additivity and independence. In the research, it was found that the variable Gender was not independent of other variables and was not additive as the two genders had different sets of variables. This would indicate that a variety of geographic areas would have to be handled by creating separate pairs of Gender Models for geographic areas. Whilst this is not impossible, it would require a very large sample in order to supply sufficient large subsets for each model.

The set of two Gender Models developed by this research may also be limited to the particular Coach Education Programme, that of UK Athletics to the year 2000. This research only used a population that had access to the last version of the first athletic programme introduced by the BAAB. The BAAB initiated programme has been substantially revised by UK Athletics with emphasis on making the new programme more accessible and more attractive by introducing a short Introductory course (Level 1 NVQ equivalent); by creating a modular structure with cross over opportunities; and by introducing more means of qualifying and of studying (UK Athletics, 2000.). The effect of this new programme is a lower entry level for athletics coaching for all those involved in athletics from the year 2000. The effect on modelling of a lower entry level is to lower the threshold value in the underlying conceptual model illustrated in Diagram 5.3.1.

There are two immediate consequences. The present research can not be replicated if the data is from samples taken post 2000, subsequent to the introduction of UK Athletics new Coach Education Programme. Comparison may be difficult of the present research with models made data arising from the new Programme.

Separate models may be required for Coach Education Programmes of other sports, especially if their entry levels differ from or where the programmes are delivered in a substantially different way to that of the original athletics programme. The key issues are whether or not the entry levels effect the decision threshold and whether or not the sport and programme effect the set of Explanatory variables. For example, just as
the two Genders appear to have different sets in this research, team sports may have different sets of variables to individual sports and sports mainly using paid coaches may have different sets to those sports with mainly unpaid coaches.

These limitations have always existed in decision research, whatever the methodology. This research only emphasises the need to recognise that gender, geographic area, entry level and activity area (in this case sport). Unless these are adequately recognised, research outcomes can not be fully compared for significant differences across these areas.

13.3. Potential Improvements to the Research

Two major improvements could be made to the research. All of these improvements have come to notice as a result of the research to-date.

- Improvement in coefficient standard errors
- Reduction in correlations between variables

These are now discussed.

The standard errors of the coefficients are too high to allow the calculated values of the Coefficients to be statistically tested, particularly in both Gender Models. As a result, meaningful tests can not be made on whether or not two coefficients within a categorical variable are significantly different; tests can not be made on the significance of coefficients from zero; and tests can not be made between values of the same coefficient between Models (e.g. between the Male and the Female Model).

Four causes exist:

- the sample size used to construct a model.
- the number of categorical explanatory variables and the number of classes within each of these categorical variables.
- dispersion.
The target sample size for this research was set at 120 on the premise that the number of variables would lie in the range of 9 to 17 variables. In this context, the number of variables refers to the number of coefficients in the model. Whilst a binary variable or a continuous variable each produces a single coefficient, a categorical variable produces \((C-1)\) coefficients where \(C\) is the number of classes in the categorical variable.

The Thesis Model is constructed from data from 112 cases and has 12 explanatory variables. But 8 of the 12 variables are categorical, each with 2 or more classes. As a consequence, there are 22 coefficients, above the maximum number planned. The Male Model is constructed from data from 62 cases and has 9 variables giving 14 coefficients. The Female Model is constructed from data from 48 cases and has 7 variables giving 10 coefficients. Both have significantly fewer cases than suggested by Tabachnick & Fidell (1997).

The need for a much larger sample is further increased by the suggestion from the research that there may be two distinct populations of people entering the coach education programme: those that enter with the prime intention of coaching at a club and those whose prime interest is to obtain a qualification to advance their career with little or no intention of immediately and regularly coaching at a club.

The SPSS programme during model building advised that the co-variance matrix for all three models was sometimes too sparse for complete calculation. This indicates that there may be too few cases to fill all the cells in the co-variance matrix being used. The more coefficients that are being calculated, the more likely this situation is to occur. As a consequence, the more categorical variables that are used, the more chance there is of sparse matrices.
Over dispersion is often caused by omission of one or more significant explanatory variables. Omission may be because one or more variables are unknown or can not be measured. It may also occur because of high correlations between potential variables allowing only some of them to be used in model building to reduce collinearity to an acceptable amount (Tolerance statistics below 0.20).

High correlation may occur because one or both of the variables are measuring, in one metric, a number of primary effects. One example in this research is Flesch Ease of Reading Score and school leaving age. This is because equation for Flesch was based upon a regression model with school age. Another probable example in this research is car availability. This variable is probably measuring socio-economic factors, area of residence, attitudes to car ownership, gender make-up of the household, age make-up of the household, work and social patterns, etc. as well as the number of cars.

The indications from this research are that one or more primary variables of significant importance to the decision process had not been identified in the published research used herein. Some indication of what these missing variables might be was picked up in the qualitative part of the interviews in this research, notably the socio-economics of the populations associated with athletics and the types and frequency of other leisure activities of the household. Other indications are that attitudes towards and perceptions of volunteering may be crucial and that these attitudes may be influenced both by culture (and ethnicity) and generation.

The set of variables used to construct the model could be improved by first conducting extensive open ended interviews with samples of parents and members aimed at identifying the factors that appear to influence the thinking, or lack of thinking, about volunteering and coaching.

The last reason for high errors is the use of data weighting. This was used in all three models. Data weighting is needed when stratified sampling is used. In the research sample, strata were formed around the categories of association (coaches, members
and parents) and the two gender. The association stratum was introduced with a target 100% sampling of the coaches as there were, relatively, few of the latter and it was felt essential to include as much data as possible about this group. 100% sampling of the entire population was not, economically, feasible. Weighting was also needed to correct for the imbalances resulting from the differing rates of conversions to interviews of the requests for interview.

Weighting can cause co-variance scarcities and estimation biases because the method of weighting is to duplicate each case by the appropriate weight to simulate the entire population. However, this method simply duplicates cells within any matrix.

Six partial solutions are available for future research, especially if greater resources are available:

i. Increase the sample size, thus giving more cases

ii. Sample a larger proportion of the population, minimising the effect of weighting

iii. Use non stratified sampling, eliminating the need for weighting

iv. Decrease the proportion of categorical variables or the classes within categorical variables or transform non-linear variables by means of a mathematical shape (e.g. square root or inverse). Any of these will decrease the relative number of coefficients.

v. Identify variables which are not highly correlated

vi. Identify more fundamental variables

Which solution or solutions are employed will be dependent upon the characteristics of the population, the interview conversion rates, the feasibility of using transforms (and the ease of interpreting and justifying any transforms) and the time and resources available.
Having identified the two major areas in which the research could be improved, the next Section looks at the measurement of adult reading.

13.4. Implications for Sport in the UK

Three major issues confront sport in the UK at present, all of which are being addressed by new initiatives from Sport UK and the National Governing Bodies.

- How to increase the number of volunteers, especially in coaching, and deploy them effectively
- How to ensure and improve the standard of training of volunteers
- How to incorporate the skills of professional administrators into a predominantly voluntary activity

The research indicates possible implications for all three of these issues. The models developed in this research may be further exploited to understand and clarify specific variables and processes involved.

The research in a restricted domain (single sport, limited geographic area) clearly indicates that the decision processes are complex and involve many variables, some acting in a non-linear fashion. Even within such a restricted domain, there were clear indications of a number of distinct populations, each operating in different ways. In the case of this research, these populations were the two genders and those who had career motives and those who had leisure motives. Such complexity invalidates the assumptions of the commonly used quantitative techniques, including t-tests, chi-square, ANOVA and linear regression.

The research presents a Sport Decision conceptual model and shows that this model can be statistically represented by a logistic regression model, allowing quantitative analysis and hypothesis testing respecting much of the complexity.
The qualitative element of the research clearly indicated a large and widespread lack of knowledge amongst both parents and athletes of the duties of a coach and how to become a qualified coach. In the case of athletics, even before a family first goes down to an athletics club, there is an assumption that there is adequate and qualified coaching available. This assumption appears to arise from a logical association with the fact that most athletics tracks are local authority facilities and often used by the local education authority.

Coupled with this lack of knowledge is a widely held perception that a coach should have practised the sport his or her self or, at least, been proficient in a similar sport. This perception may be reinforced by feedback from athletes, especially those in their late teens and early twenties. As with most sports, participation by women in athletics after leaving school is, relatively, new. As a consequence, few present day mothers have competed. This may be one reason why there are fewer female coaches than male.

The research indicates that voluntary coaches mainly come from parents with children, especially girls, who are already members of a club. Parents volunteer to become coaches to, directly or indirectly, help their own children because they see participation in sport as beneficial to the social and health education of the children. It is not an altruistic act in the first instance. They volunteer because they perceive that there is a shortage of coaching assistance either from observation or from feedback from their children or by being asked for help by an existing coach.

The biggest deterrent to becoming a coach appears to be the perception that a coach has to be regularly at the club two or three times every week and spends a significant amount of time both planning and operating. This perception of significant and regular commitment arises from the standard practise whereby a coach services all the coaching needs of a fixed group of athletes who look to this specific coach to organise every session. Time commitment is seen as a major problem by parents in an
age of work pressure, two parent working and an ever widening range of physical leisure activities both for children and, increasingly, for adults.

The stability arising from the continuity and regularity of a coaching group is recognised by coaches, parents and young people as being an essential feature of a successful and happy club. A group can not be artificially created. It exists through a leader and the acceptance by the group of the abilities of that leader, natural abilities enhanced by the technical training the leader has received. Whilst the frequent presence of the leader is probably essential for the group to survive, there are no reasons why this presence has to be at every session. Groups continue during the leaders absence on holiday and for illness and when the leader is unable to get away from work. Such absences are usually covered by the leader issuing written schedules or by instructing another coach or person what to do in their absence.

There is also an increasing concern about apparent growing litigation relating to child abuse and to injury. Interviewees frequently said that they did not wish to take on responsibilities that exposed them to such dangers. Ironically, very few parents seemed to be concerned about these issues relating to their own children and virtually none of those interviewed had formally ensured that the volunteers at the club had been vetted in any way or were formally qualified. Again, the assumption was that this was satisfactory because the facility was provided by a local authority.

The other obvious source of new coaches is existing athletes. Whilst many of them expressed some desire to "give something back to sport", most of them seem unlikely to coach on a regular basis whilst they can compete. Coaching is seen as a replacement for competition as a means of retaining association with the club, probably for social reasons. Again, entry to coaching by ex athletes is probably not for altruistic reasons. Two recent developments in athletics have greatly increased the ability of an individual to remain competing into their 30's, 40's and 50's. These are the emergence of road running in the 1980's and the more recent increase in veteran athletics of the 1990's. They are already impacting on the recruitment of coaches and
this impact is likely to increase over time. Any apparent increase in young people entering the Coach Education Programme may be misleading. The interviews in this research clearly showed that most of them entered to enhance their careers and to boost their CV’s for applications to Higher and Further education. Very few of them subsequently coach regularly at clubs.

Due to the lack of general knowledge about Coach Education, it was only possible to gain insight into revalidation and upgrading of qualifications from existing coaches.

Whilst the concept of revalidation was, fairly generally, agreed, the method used would seem to be crucial to gaining general acceptance. This seems to arise from how a coach develops in a club. Although this development was not a subject for the research, some aspects did emerge during the interviews. Coaches who practised for several years appear to be those who are able and willing to commit many hours every week on a regular basis, adjusting their family, social and work lives accordingly. Their coaching tends to be centred on a group of athletes, often referred to as ‘my athletes’. A great loyalty or dependency seems to be a feature of this organisation. It becomes an integral and important part of the life style of the coach, often a dominant part of his or her leisure. Time away from this group is viewed as time lost. Possibly there is a fear of losing control of or weakening membership of the group if too much time is spent away. This may be reinforced by experiences of other coaches ‘stealing’ athletes. This coaching group integrity seemed to be an essential component of feeling competent to coach. Going away from the group to be revalidated was both threatening and worrying. The time lost away from the group was resented. The requirement to demonstrate coaching ability with unknown people was feared.

Similar concerns of being away from the group surrounded thoughts about upgrading the level of coaching qualification. In addition, there were perceptions that the higher levels of coaching qualification placed more and more emphasis on the theoretical rather than the practical aspects. Many respondents stated that it was the practical aspects of coaching which gave them pleasure. Many coaches could see no practical
improvement to their relationships with ‘their athletes’ by upgrading their qualification. Part of the reason behind these attitudes may stem from the fact that most Level 1 coaches deal with young children at or soon after their introduction to athletics. Much of their time is spent on fundamentals, less in ability development, especially at an individual rather than a group level. This follows from how and why most parent enter coaching.

Of much greater concern is an often expressed need to be able to better control young people, especially boys. The problems are perceived as getting worse with growing bad behaviour and growing lack of interest from the parents. Respondents frequently complained about becoming ‘child minders’. The problem was also identified by some parents who were teachers as they observed these voluntary coaches. The suggestion that teaching fundamentals of handling children be introduced into Level 1 was well received by several respondents and supported by the teacher parents.

Another factor that may be changing at present is the attitude towards volunteering, especially where volunteering is associated with not being paid. This interviews found that the present generation of coaches tended to hold strong views on the desirability of people giving their time and for voluntary coaching to be undertaken without payment. This was particularly true of the female respondents. Yet many of the younger members and parents held views that can be summed up as “You get what you pay for”. They either held neutral views about paying coaches or felt that payment would increase recruitment and, probably, improve standards. Yet most of those holding these views could not see there being sufficient money available to pay coaches and that payment would disadvantage many children. These findings may reflect the decline in volunteering amongst the younger age groups reported in the 1991 National Survey of Volunteering (Lynn, P, and Smith, J D, 1991.).

The interviews indicate that the introduction of paid professionals into administration and coaching in athletics must be handled with great care if beneficial and willing cooperation is to be obtained with the huge numbers of unpaid volunteers in the grass

RESEARCH OUTCOMES
root clubs. Whilst the abilities of a professional are appreciated, the danger lies in them seeing to be attempting to direct the decisions and actions of the volunteers. They can only be effective if they act in advisory capacities at the request of the clubs. Many of the present tensions seem to be occurring because the volunteers feel that there is a lack of regular communication. The communication expected will not be entirely satisfied by post and Internet. What is expected is the willingness, even desire, for the professionals to spend time observing and listening at the clubs.

These observations indicate that enhanced recruitment to voluntary coaching is possible if modification can be achieved to the attitudes and perceptions that are currently acting as constraints. Insight into how suitable modifications can be made can be obtained from psychological models of change. An integrated model of change was proposed by James Prochaska (1982). He puts forward a schema for the progress of change as a revolving system having five quadrants: Contemplation, Determination, Action, Maintenance and Relapse. The schema is entered in the Contemplation quadrant. There are three exit points. The determination process may result in a decision to make no changes, e.g. not to become a coach. The maintenance process may result in a decision to change use of free time, e.g. to become a coach. The relapse process may result in a reinforcement of original patterns, e.g. to cease coaching.

The most appropriate means of influencing the cycle depends upon the process within the cycle that an individual is in. At any one time, there will be many potential recruits to coaching in each of the five processes. This indicates that it is necessary to have a variety of strategies for change in place all the time. This contrasts with the existing state in athletics which appears to rely heavily on verbal encouragement to coach from existing coaches and club officials and on the observations and consequent reactions of parents. The interviews in this research indicate that perceptions play an important part, usually constraining, in the determinations of parents.
Prochaska's (1982) integrative model groups change methods into five strategy types, each type having methods working at the experiential level or at the environmental level. Their work gives examples of the use of appropriate types of intervention in each of the processes within the cycle of change. The use of the Sport Model and of the methodology put forward in this research can be used to indicate the variables most influential in the decision process and the setting of the motivational threshold. The combination of the two approaches, the psychological and the mathematical, may be useful in formulating new strategies for the recruitment and retention of voluntary coaches and for improved rates of qualification upgrading.

Having examined the implications of the research for a specific form of adult education and volunteering, coaching, the last Section examines the implications for adult education in general.

13.6. Implications for Adult Education in the UK

In the UK decisions on entry to Primary and Secondary education are largely confined to a choice of establishment. The making of a decision is mandatory by Law. The choice is usually dominated by residency.

In contrast, choices and decisions in Adult Education are mainly voluntary. In this respect, these decisions are similar to the decision to enter a voluntary Coach Education Programme. Another factor that both types of decision have in common is that the decisions both tend to have an element of perception of the contents of a course and of the outcome. These commonalities suggest that the conceptual model and the practical logistic regression model developed in this research could be adapted to decisions on entering specific Life Long Learning programmes. The benefits of using such techniques are likely to be a better understanding of how prospective students make decisions and the relative importance of the significant factors associated with the decision, both in terms of incentives and disincentives. Such
knowledge could be utilised both in the marketing of the programmes and in their construction and contents.

What the models developed in this research indicate are that such decisions are probabilistic and are associated with several variables, some of which may interact with each other. Also, variables can not be assumed to act in a linear fashion. As a consequence, analysis using simple, univariate techniques such as Chi Square and Linear Regression may not be appropriate and, if used, may give misleading results.

In addition, the research indicates that the set of significant variables related to a specific programme may be specific to a population. Or, at least, the relative significance of individual variables may vary with population. For example, in this research it was found that men and women were best considered as separate populations. There was also some indication that people primarily motivated by career may constitute a different population to those motivated by desires not directly connected with their careers.
REFERENCES


REFERENCES


Odeh, R E, and Fox, M, 1975 Sample size choice: Charts for experiments with linear models. NY: Marcel Dekker


REFERENCES


APPENDIX A

Initial Contact Letter
(University Headed Paper)

Dear xxxxxx

I am doing research in the Department of Education at the University of Surrey. I am looking at parents, participants and coaches in voluntary, amateur sport clubs. The research is independent of government, companies or the Sports Council.

XXX Club are collaborating in this research and supplied me with your name and address as a Member.

By individual interview, I am collecting the experiences and views of a sample of members and parents. All answers will be strictly confidential to myself. Research reports will not give names and will be written so that individuals can not be identified.

I hope that you will allow me to interview you. Enclosed is a reply with a prepaid envelope. Alternatively, you can contact me by e-mail or telephone. If I do not hear from you, I will telephone you shortly to arrange a suitable time, day and place. I will understand if you do not wish to participate. If you do, you can withdraw at any time. I prefer to come to your house or office (day or evening, weekday or weekend), or I can meet you at XXX.

You may wish to check on me from the Chairman of XXX Club, XXX (telephone number) or from Simon Ingyon, Surrey Sports Development Unit (01483 532614) or from my Head of Department, Professor Karen Evans (01483 300800).

Sport is undergoing major changes. One hope is that this independent research from the 'grass roots' may have some influence. I do hope that you can participate and make your views known. I apologise if you object to this contact.

Yours sincerely

XXX XXX
APPENDIX B

APPENDIX B - REPLY SLIP

Reply slip with Contact letter

Please return to: (Researcher, address)

From

I have read your letter and (Cross out as applicable)

1. I DO NOT wish to participate in the research. Please DO NOT contact me again.

2. In principle, I am willing to participate in the research. The most convenient times, days, and place to meet are:
   
   Times:
   
   Days:
   
   Places:

3. The person you wrote to no longer lives here.
   Their present address is:
APPENDIX C

COACHING QUESTIONNAIRE & SHOW CARDS - COACHES

Respondent Number

Show Card Type

QUESTION 1.

Respondent is:

Male □ Female □

QUESTION 2.

Respondent is a member of:

A Club □ B club □

“This interview is with Respondent Number .......... It was recorded on .......... at ..........”

“The research I am doing is about people and sport. It is not for government or for any company. Your name will not appear in anything written, nor will you be identifiable. All the information which you give me will be confidential to myself and kept in a secure place. You can terminate this interview at any time. You can ask questions at any time. To maintain the confidentiality, what name shall I call you by for this recording?

.................................................................

QUESTION 3.

“Am I correct in saying that you are, or have been, a qualified athletics coach?”

(Definition: hold or have held a Level 1 or Assistant Club Coach Award from the British Amateur Athletics Board, BAAB, or the British Athletics Federation, BAF, or UK Athletics)

Correct □ Incorrect □

If “Incorrect”, change Questionnaire
If “Correct”,

QUESTION 4.

“Would you kindly tell me your year of birth?”

(The main reason for wanting this information is to allow for changes in attitudes and views over the generations)

Year □
QUESTION 5.
“In which year did you first register to become a Level 1 Coach or Assistant Club Coach? The approximate year will do if you are not sure of the exact year.”
Year of first Registration (YR)

QUESTION 6.
“What is the highest level of athletics coaching qualification you have passed?”
Level 1 (Assistant Club Coach) ☐ Level 2 (Club Coach) ☐ Senior Coach ☐

QUESTION 7.
“Do you still hold a valid athletics coaching award?”
Yes ☐ No ☐
If “Yes” go to Question 9

QUESTION 8.
“In which year did the validity expire?”
Year of Expiry

QUESTION 9.
“Is or was your coaching mainly at a voluntary athletics club, as opposed to where you work or at a club or association not specifically for athletics?”
Yes ☐ No ☐
If “Yes”, go to Question 13.

QUESTION 10.
“Over the past 12 months, were you normally coaching athletics at least once a week?”
Yes ☐ No ☐
If “Yes”, go to Question 13.

QUESTION 11.
“In what year did you last normally coach athletics at least once a week?”
Year

APPENDIX C - QUESTIONNAIRE
Respondent Number

QUESTION 12.
"What was or were the main reason or reasons for you not normally coaching athletics at least once a week?"

Reasons

Go to Question 14

QUESTION 13.
"Over the past 12 months, how many days a week did you normally coach athletics?"

One  Two  Three  Four  Five or more

QUESTION 14.
"Excluding travel, over the past 12 months, please estimate how many hours per week you normally coached athletics."

Hours per week

QUESTION 15.
"Since you first registered to become an athletics coach, i.e. since ......., have you ever held another or other official post or posts in an athletics club?"

<table>
<thead>
<tr>
<th>Official</th>
<th>Manager</th>
<th>Committee</th>
<th>Other</th>
<th>None</th>
</tr>
</thead>
</table>

"Many of the following questions refer to you AT THIS DATE, when you first registered to become a coach. Please estimate if you can not fully remember."

QUESTION 16.
"When you first registered to become an athletics coach, were you living at your present address?"

Present address (Place, Street)

Yes  No

If “Yes”, Go to Question 18.
If “No”,

QUESTION 17.
"When you first registered to become an athletics coach, where were you living (Place & Street)?"

Original address

APPENDIX C - QUESTIONNAIRE
QUESTION 18.

“When you first registered to become an athletics coach, on a weekday evening, how long did it normally take you to get to the athletics track?”

Number of Minutes

QUESTION 19.

“When you first registered to become an athletics coach, outside work or study hours you had then, how many cars and or vans did you normally have available to you?”

(No. > 4)

None □ One □ Two □ Three □ 4 or more □

QUESTION 20.

“During the past 12 months prior to you registering to become an athletics coach, did your work or study involve:

• shifts?
• or irregular hours (including irregular overtime)?
• or frequently (at least twice a month) staying nights away from home?”

Yes □ No □

QUESTION 21.

“When you first registered to become an athletics coach, what was the earliest time on a Tuesday or a Thursday evening that you were usually free from business, commuting, education or normal household duties?”

[If shift work, when shift did NOT include evening. Allow for sleeping.]

[If work away from home, only when at home]

Hour □ Minutes □ AM □ PM □

Decimal 24 Hour

QUESTION 22.

“When you first registered to become an athletics coach, were there other people living with you and dependent upon you?”

[Dependent = Children under 18; spouse or partner; other adults dependent because of age, health, injury or other disability]

Yes □ No □

If “No”, Go to Question 25

APPENDIX C - QUESTIONNAIRE
Respondent Number

QUESTION 23.

"I am interested in how many of these dependants that you had at that point in time, whether they were male or female and whether they were regularly involved with an athletics club."
[Regularly involved = usually took part or helped at least once a week during the season]

"How many boys UNDER 18 were living with you?"
Total number of BOYS UNDER 18

"How many of these boys were involved then with an athletics club (not school)?"
Total number of BOYS UNDER 18 INVOLVED in athletics

"How many girls UNDER 18 were living with you at that time?"
Total number of GIRLS UNDER 18

"How many of these girls were involved then with an athletics club (not school)?"
Total number of GIRLS UNDER 18 INVOLVED in athletics

"How many adults (18 and over), including yourself, were living with you and dependent upon you?" [include spouse or partner; adults dependent because of age, health or infirmity]
Total number of ADULTS

"How many of these adults were male?"
Total number of MALE ADULTS

"How many of these male adults were involved with an athletics club, at that time?"
Total number of MALE ADULTS INVOLVED in athletics

"How many of the female adults were involved with an athletics club at that time?"
Total number of FEMALE ADULTS INVOLVED in athletics

QUESTION 24

"In the 12 months prior to you first registering to become an athletics coach were ANY of these dependants been involved in ANY OTHER SPORT?"
Yes (0) No (1)
QUESTION 25.

"Who was the first person in your household to become a member of an athletics club (not school)?"
Self □ Partner □ Child □ Parent □ Other □

"In which year did you/he/she become a member (of XX club)?"
Year of First Membership FOR DEPENDENT □

QUESTION 26.

"In which year did you FIRST become a member of any athletics club (not school)?"
Year of First Membership FOR RESPONDENT □

QUESTION 27.

"Was this your present club (XX) or another?"
Present Club □ Another Club □

If “PRESENT”, go to Question 29

QUESTION 28.

"Did you ever compete &/or train with this or any other athletics club (other than school)?"
Yes □ No □
(0) (1)

QUESTION 29.

"At the time of registration, did your work or study involve you in any form of sport?"
(This includes being a Teacher involved with sport; working in a Sports Club or Leisure Centre; being a professional coach or trainer or sports person; a Local Authority Officer dealing with sport; selling or marketing mainly sports clothing or equipment.)
Yes □ No □
(0%)

If No, go to Question 31

QUESTION 30.

"At time of registration, what percentage of your work time (or study) involved or related to any form sport?"
Percentage □
Respondent Number

QUESTION 31.

“In which year did you become a member of your present club (XX)?”

Year of Present Membership FOR RESPONDENT

QUESTION 32.

“When you first registered to become an athletics coach, did you or did you normally compete &/or train at the athletics club?”

Normally: if not injured, ill, business, away

Yes ☐ No ☐

QUESTION 33.

“When you first registered to become an athletics coach, were you involved in this Club, other than in competing &/or training?”

(Committee, Official, Team Manager)

Yes ☐ No ☐

QUESTION 34.

“In the past 12 months prior to first registering as a coach, how were you involved in the Club?”

Tick as many as apply

Compete/Train ☐ Committee ☐ Official ☐ Manager ☐ Coach ☐

(Committee, Official, Team Manager)

Parent only ☐ Other ☐

QUESTION 35.

“During those 12 months prior to first registering to become an athletics coach, were you regularly involved, not just as a spectator, in any other sport, other than athletics?”

[Regularly involved = usually participating or helping at least once a week during the season]

Yes ☐ No ☐

(0) (1)

QUESTION 36.

“Since leaving school and up to when you first registered to become an athletics coach, for how many years had you or did you regularly compete in any sport for any club, including athletics? Please make an estimate.”

Number of Years

[Regularly competed = competed at least once in a season. Include years when injured or abroad if you returned to competition after the injury or being abroad]
Respondent Number

QUESTION 37.
"During the 12 months prior to you first registering to become an athletics coach, had you competed in athletics, or would you have done so if you had not been injured or abroad?"

Yes □  No □

QUESTION 38.
"Up to when you first registered to become a coach, had you ever competed or did you intend to compete as a Veteran in athletics?"

Yes □  No □

QUESTION 39.
"When you first registered to become an athletics coach, would you say that your MAIN interest in athletics was Road Running?"

Yes □  No □

SHOW CARD ‘A’

QUESTION 40.
"In the 12 months prior to first registering to become an athletics coach, how often were you helping with and or competing in any formal organised sport or exercise (other than recreational walking)?"

Yes □  No □

If “NO”, go to Question 43

QUESTION 41.
"Is this answer different NOW?"

Yes □  No □

QUESTION 42.
"Would you please mark the scale again to show how you might have felt then."
“At the present time, to what extent do you agree or disagree with the following statement: People should volunteer some of their time to people elsewhere.”

SHOW CARD ‘B’ (second copy)

QUESTION 43.

“Do you think that this answer might have been different at the time which you first registered to become an athletics coach?”

Yes [ ] No [ ]

If “NO”, go to Question 46

QUESTION 44.

“Would you please mark the scale again to show how you might have felt then.”

SHOW CARD ‘B’ (second copy)

QUESTION 45.

“When you first registered to become an athletics coach, did you usually have a paid job or were you studying?”

Paid Job [ ] Studying [ ] Neither [ ]

(0%)

If Neither, go to Question 49

QUESTION 46.

“At the time of registration, did your work or study involve you in any form of sport?”

[This includes being a Teacher involved with sport; working in a Sports Club or Leisure Centre; being a professional coach or trainer or sports person; a Local Authority Officer dealing with sport; selling or marketing mainly sports clothing or equipment.]

Yes [ ] No [ ]

(0%)

If No, go to Question 49

APPENDIX C - QUESTIONNAIRE
QUESTION 48.

"At time of registration, what percentage of your work time (or study) involved or related to any form sport?"
Percentage  

QUESTION 49.

"What is your present situation?"
Retired  Paid Job  Studying  Other  

If “RETIRED” or “OTHER”, go to Question 52

QUESTION 50.

"Does your present work or study involve you in any form of sport?"
[This includes being a Teacher involved with sport; working in a Sports Club or Leisure Centre; being a professional coach or trainer or sports person; a Local Authority Officer dealing with sport; selling or marketing mainly sports clothing or equipment.]
Yes  No  (0%)

If No, go to Question 52

QUESTION 51.

"What percentage of your work time (or study) involves or is related to any form sport?"
Percentage  

QUESTION 52.

"I am interested in how much full or part time education you had up to the time you first registered to become an athletics coach."

"At what age did you leave school?"
Age  

Years after 15  

QUESTION 53.

"Did you go or are you at a College or University, prior to first registering to become an athletics coach?"
Yes  No  

APPENDIX C - QUESTIONNAIRE
Respondent Number

QUESTION 54.

“How many years did you spend at College and or University, up to the time when you first registered to become an athletics coach? An estimate will be OK.”

(Ask for each, number of days per week. 1 day per week = 0.2 years per year; 1 evening per week = 0.1 year per year)

Number of Years

Total Number of Years of Education after 15

QUESTION 55.

“Other than for your job, at the present time, do you read a newspaper, a magazine, a journal or a book at least once a month?”

   Yes [ ]   No [ ]

FS = 100

If “No”, Go to Question 62.

QUESTION 56.

“At the present time, do you usually read a local or national newspaper at least once a week?”

   Yes [ ]   No [ ]

If “No”, Go to Question 59.

QUESTION 57.

“Which newspaper do you most frequently read?”

   Name of Newspaper

QUESTION 58.

“Tick all the subjects which you read in this newspaper.”
Respondent Number  

SHOW CARD 'C'

QUESTION 59.

"At the present time, do you usually read a magazine or journal at least once a month?"

Yes  
No  

If "No", Go to Question 62.

QUESTION 60.

"Which magazine or journal do you most frequently read?"

Name of Magazine or Journal

QUESTION 61.

"Tick which type of articles you read in newspapers, magazines, journals and books. You may tick one or both"

SHOW CARD 'D'

QUESTION 62.

"At the present time, to what extent do you agree or disagree with the following statement?"

"Coaches at local sports clubs, generally, should be paid?"

SHOW CARD 'E'

QUESTION 63.

"Do you think that this answer might have been different at the time which you first registered to become an athletics coach?"

Yes  
No  

If "NO", go to Question 65.

QUESTION 64.

"Would you please mark the scale again to show how you might have felt then."

SHOW CARD 'E' (second copy)
QUESTION 65.
“How did you become associated with an athletics Club (not school) PRIOR to you first registering to become a coach?”

QUESTION 66.
“How did you become an athletics coach? Was there a single factor or occurrence that finally made you decide to register to become a coach?”

QUESTION 67.
“Had you been a Club (not school) athlete BEFORE you first registered to become a coach?”

QUESTION 68.
“How important is it that the club provides coaching?”

QUESTION 69.
“Would you have become associated with the Club if it had not provided coaching?”

QUESTION 70.
“Do you think that there could be a better system of providing athletic coaching other than the present system using club members?”

QUESTION 71.
“In your opinion, why do people take up athletics coaching, now and when you first registered?”

QUESTION 72.
“Do you think that this has changed since you first registered?”

QUESTION 73.
“What significant changes do you think that there have been and are being made in athletics coaching since you started coaching? Comment on any benefits and or disadvantages which you see these changes have made.

QUESTION 74.
“If you were at the same point in your life as when you started coaching, do you think that you would or would not enter the coaching scheme NOW? Please comment on your reply.

QUESTION 75.
“Are you aware of the changes to the Coach Education system presented at the 1999 Athletics Congress in Manchester in November?”

Yes [ ] No [ ]

If “No” go to Question 80.
QUESTION 76

"Have you read the Report of the Review Group?"

Yes □ No □

QUESTION 77

“What changes do you welcome and why?"

QUESTION 78

“What changes do you not welcome and why?"

QUESTION 79

“What other changes would you like to have seen in the Report and why?"

“Thank you for your help in this enquiry. This ends the interview with ..............

QUESTION 80

“What changes would you like to have been in the Report and why?”

“Thank you for your help in this enquiry. This ends the interview with ..............

(Note: variations of this questionnaire were used for members (not coaches) and parents. Several questions on this sample questionnaire are specific to coaches)

(Note: Items in italics were as spoken by the researcher to the respondent in the interview. Items in square brackets were prompts or definitions, used if needed. Items following a question and answer are directions to next question.)
PERSONAL INVOLVEMENT IN SPORT & EXERCISE

"In the past 12 months, how often did you help with and or compete in any formal organised sport or exercise (other than recreational walking)?"

Place ONE mark on the scale.

Once a week or more (1.0)
Fairly regularly, but not every week (2.0)
Irregularly, at least once a month (3.0)
Very infrequently, not usually every month (4.0)
Not at all (5.0)
SHOW CARD 'B'

VOLUNTEERING

"To what extent do you agree or disagree with the following statement:

People should volunteer some of their time to people elsewhere."

Make ONE mark on the scale.

Strongly Disagree (1.0)

Disagree (2.0)

Neither Disagree nor Agreed. Didn’t know (3.0)

Agree (4.0)

Strongly Agree (5.0)
SHOW CARD ‘C’

NEWSPAPER READING

“Tick all the subjects which you read in this newspaper.”

You may tick as many as applied

<table>
<thead>
<tr>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sport</td>
</tr>
<tr>
<td>Government &amp; Politics</td>
</tr>
<tr>
<td>UK &amp; local News &amp; stories</td>
</tr>
<tr>
<td>Foreign News &amp; stories</td>
</tr>
<tr>
<td>Travel</td>
</tr>
<tr>
<td>Fashion</td>
</tr>
<tr>
<td>Business &amp; Finance</td>
</tr>
<tr>
<td>Theatre, Arts &amp; Books</td>
</tr>
<tr>
<td>Motoring</td>
</tr>
<tr>
<td>Science &amp; technology</td>
</tr>
</tbody>
</table>
SHOW CARD ‘D’

TYPES OF READING

“Tick which type of articles you read in newspapers, magazines, journals and books. You may tick one or both”

Mainly facts and details

About people & what they said and did
PAYMENT OF COACHES

"To what extent do you agree or disagree with the following statement/"

"Coaches at local sports clubs, generally, should be paid?"

Make ONE mark on the scale

Strongly Agree (1.0)

Agree (2.0)

Neither Disagree nor Agree. ‘Don’t know’ (3.0)

Disagree (4.0)

Strongly Disagree (5.0)
APPENDIX D

A Position Paper of the Volunteerism Commission
World Leisure & Recreation Association

1. Volunteering, and all other leisure activities, must be freely chosen to be satisfying and rewarding; if they are in some way coerced, they are cannot really be seen as volunteering.

2. Volunteering is more than just enjoyable, it is also rewarding. Nevertheless, there are some disagreeable features in all volunteer roles (i.e., the requirement to maintain an inflexible schedule), and there is a degree of obligation similar to, but perhaps not as strict as, that which occurs at work.

3. There are three motives for volunteering: altruism, collective interest and self-interest, of which the latter two most clearly reflect the leisure aspect of true volunteering.

4. Rarely are leisure studies and voluntarism/citizen participation discussed together, despite their theoretical compatibility.

5. Mainstream volunteering is characterized by an exceptional sense of obligation (which is the motivational force behind their participation). This is different from marginal volunteering which feels, to the person involved, as though it has a greater degree of disagreeable, inflexible coercion. Such marginal volunteering occurs when one feels that the work s/he is contributing freely is being exploited as unpaid labor. Obviously, the leisure component of marginal volunteering is diluted, if this volitional (or freely motivated) element is lessened.

6. In marginal volunteering, volunteers will tend to feel less motivated and obligated than mainstream volunteers. Therefore, those who recruit volunteers must realize this difference, and understand that they may not be able to count on the marginal volunteers in the same way as they do upon mainstream volunteers.

7. In the Information Age, there is a decline in government funding of social services, and responsibility for such services is shifted to voluntary organizations. If globalization also lessens the power of ordinary citizens, then volunteerism becomes increasingly important, and people can become empowered, through volunteerism, to participate more fully in community building. If society is increasingly dependent upon the goodwill of such volunteers, it is thus important to understand who they are, why they volunteer and what they are able to contribute. Communities must realize that true volunteering is a leisure activity, while marginal volunteering does not qualify as leisure.

R.A. Stebbins, December 2000.