Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author’s right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. http://researchspace.auckland.ac.nz/feedback

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form.
The Auckland Blood Pressure Control Study

A randomised controlled trial of physical activity and salt restriction in persons being treated with medication for hypertension

by Bruce Arroll

A thesis submitted for the degree of Doctor of Philosophy at University of Auckland Auckland, New Zealand

October 1992
### Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Abstract</strong></td>
<td>ix</td>
</tr>
<tr>
<td></td>
<td><strong>Acknowledgments</strong></td>
<td>xiii</td>
</tr>
<tr>
<td></td>
<td><strong>List of Figures and Tables</strong></td>
<td>xv</td>
</tr>
<tr>
<td>1</td>
<td><strong>Chapter 1: Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1.1</td>
<td>General Introduction</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Hypertension, salt intake, physical activity and coronary heart disease</td>
<td>5</td>
</tr>
<tr>
<td>1.1.3</td>
<td>Study overview</td>
<td>9</td>
</tr>
<tr>
<td>1.1.4</td>
<td>Aims</td>
<td>10</td>
</tr>
<tr>
<td>1.1.5</td>
<td>General aims: Auckland Blood Pressure Control Study</td>
<td>10</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Basic principles in a randomised controlled trial</td>
<td>11</td>
</tr>
<tr>
<td>1.2.2</td>
<td>General feature of a randomised controlled trial</td>
<td>13</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Ethical issues in a randomised controlled trial</td>
<td>21</td>
</tr>
<tr>
<td>1.2.4</td>
<td>Validity</td>
<td>24</td>
</tr>
</tbody>
</table>
### Chapter 2: Literature review: Auckland Heart Study Validation Project

2.1.1 Introduction | 29
2.1.2 Overview of physical activity literature | 29
2.1.3 Validation of physical activity studies | 33
2.1.4 Seven day recall and seven day diaries | 34
2.1.5 Studies with fitness as a gold standard | 36
2.1.6 Studies of motion activity sensors | 42
2.1.7 Validation of "simple questions" | 45
2.1.8 A study of multiple methods of measuring physical activity | 48
2.1.9 Repeatability | 49
2.1.10 Conclusions | 50

### Chapter 3: Methods: Auckland Heart Study Validation Project

3.1.1 Introduction | 58
3.1.2 The Auckland Heart Study | 58
3.1.3 Auckland Heart Study: Three month physical activity recall | 60
3.1.4 Auckland Heart Study: Three month food frequency recall | 61
3.1.5 Auckland Heart Study Validation Project: Seven day Self Report Diary | 62
3.1.6 Methods | 63
3.1.7 Statistical methods | 67
3.1.8 Rationale for choosing a one week diary | 69
3.1.9 Summary | 72
Chapter 4. Results of the Auckland Heart Study Validation Project.

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.1</td>
<td>Introduction</td>
<td>75</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Results</td>
<td>75</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Hours spent in work, leisure and home activities</td>
<td>76</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Measures of correlation and agreement</td>
<td>78</td>
</tr>
</tbody>
</table>

Chapter 5: Discussion

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1.1</td>
<td>Introduction</td>
<td>88</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Discussion</td>
<td>88</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Comparison with other studies</td>
<td>92</td>
</tr>
</tbody>
</table>

Chapter 6: Background and literature review for the Auckland Blood Pressure Control Study

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1.1</td>
<td>Introduction</td>
<td>97</td>
</tr>
<tr>
<td>6.1.2</td>
<td>Overview of the literature for physical activity and hypertension</td>
<td>97</td>
</tr>
<tr>
<td>6.1.3</td>
<td>Methods for review of the literature for physical activity and hypertension</td>
<td>98</td>
</tr>
<tr>
<td>6.1.4</td>
<td>The Seals and Hagberg review</td>
<td>100</td>
</tr>
<tr>
<td>6.1.5</td>
<td>Results of the current review of the literature for physical activity and hypertension</td>
<td>105</td>
</tr>
<tr>
<td>6.1.6</td>
<td>Discussion</td>
<td>117</td>
</tr>
<tr>
<td>6.1.7</td>
<td>Conclusion</td>
<td>121</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>6.2 Salt and hypertension</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>6.2.1 Introduction</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>6.2.2 Experimental studies</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>6.2.3 Recent articles on salt restriction</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>6.2.3 The Intersalt Study: an ecological study.</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>6.2.3 Community studies</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>6.2.4 General Practice studies</td>
<td>133</td>
<td></td>
</tr>
<tr>
<td>6.2.5 Medication and blood pressure</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>6.2.6 A Review of the literature</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>6.2.7 Conclusion</td>
<td>136</td>
<td></td>
</tr>
</tbody>
</table>

**Chapter 7: Methods**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1.1 Introduction</td>
<td>149</td>
</tr>
<tr>
<td>7.1.2 Participants</td>
<td>149</td>
</tr>
<tr>
<td>7.1.3 Randomisation</td>
<td>151</td>
</tr>
<tr>
<td>7.1.4 Interventions</td>
<td>155</td>
</tr>
<tr>
<td>7.1.5 Data Collection</td>
<td>158</td>
</tr>
<tr>
<td>7.1.6 Outcome measures</td>
<td>161</td>
</tr>
<tr>
<td>7.1.7 Scoring system for medication</td>
<td>162</td>
</tr>
<tr>
<td>7.1.8 Sources of Bias:</td>
<td>165</td>
</tr>
<tr>
<td>7.1.9 Coding and analysis</td>
<td>166</td>
</tr>
<tr>
<td>7.1.10 Statistical Power</td>
<td>167</td>
</tr>
<tr>
<td>7.2.1 Summary</td>
<td>168</td>
</tr>
</tbody>
</table>

**Chapter 8: Results from Auckland Blood Pressure Control Study**

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1.1 Introduction</td>
<td>170</td>
</tr>
<tr>
<td>8.1.2 Results</td>
<td>171</td>
</tr>
</tbody>
</table>
Tables 8.1-8.9: Baseline measurements 171

Tables 8.10-8.15: Analysis of the two exercising subgroups versus the two non-exercising subgroups. 177

Tables 8.15-8.19: Univariate analysis and multivariate analysis of two salt restricting subgroups versus the two non salt restricting subgroups. 181

Tables 8.20-8.28: Univariate analysis and multivariate analysis by the four subgroups 186

Tables 8.30-8.35 191

Tables 8.36-8.47: Age and gender subgroups 197

Tables 8.48-8.52: Medication changes and non-pharmacological therapies, summary of differences between group A and groups B and C. 205
Chapter 9 Discussion: Auckland Blood Pressure Control Study

9.1.1 Introduction 264
9.1.2 Overview 264
9.1.3 Strengths of the study 265
9.1.4 Weaknesses of the study 266
9.2.1 General findings 273
9.2.2 Patterns of blood pressure changes 274
9.2.3 Physical activity 275
9.2.4 Salt restriction 278
9.2.5 Two interventions versus one intervention 281
9.2.6 Failure to achieve statistical significance at six month interview 283
9.2.7 Weight and alcohol 284
9.2.8 Lipids and pulse 285
9.2.9 Age and gender 286
9.3.1 Implications for clinicians 287
9.3.2 Public health implications 293
9.3.3 Further research 296

Chapter 10 Conclusion:

10.1.1 Introduction 303
10.1.2 Auckland Heart Study Validation Project 303
10.1.3 Auckland Blood Pressure Control Study 304
References

Appendix A  Validation Diary and physical activity questionnaire

Appendix B  Food Frequency Questionnaire

Appendix C  Auckland Blood Pressure Control Study Questionnaires

Appendix D  Participant information sheets

Appendix E  Salt encyclopedia

Appendix F  Papers published as part of the Auckland Blood Pressure Control Study
Abstract

Introduction

This thesis describes two studies: The Auckland Heart Study Validation Project which validated the physical activity questionnaire from an earlier case-control study (known as the Auckland Heart Study); The Auckland Blood Pressure Control Study which was a factorial design randomised controlled trial of physical activity and salt restriction as a means of further lowering blood pressure in a community sample of patients treated with anti-hypertensive medication. This study used the physical activity questionnaire validated in the Auckland Heart Study Validation Project.

Coronary heart disease is the leading cause of death and a major cause of morbidity in New Zealand. The Auckland Heart Study was a case-control study examining coronary heart disease and its risk factors and was conducted in Auckland from 1986 to 1988 (Jackson, 1989). This study incorporated a three month recall questionnaire on physical activity. The results showed that a high level of physical activity was associated with a low rate of coronary heart disease. For example the odds ratio for those with high levels of moderate leisure time physical activity was 0.78 for men and 0.39 for women. To assess the validity of the physical activity questionnaire, the Auckland Heart Study Validation Project was conducted in 1988. Both physical inactivity and hypertension are risk factors for coronary heart disease and physical activity is
known to lower blood pressure. The Auckland Blood Pressure Control study was conducted during 1989-90 in order to assess the effectiveness of physical activity as a means of lowering blood pressure in a community setting. The physical activity questionnaire used in the Auckland Blood Pressure Control study was the same one validated in the Auckland Heart Study Validation Project. This thesis describes both the Auckland Heart Study Validation Project and the Auckland Blood Pressure Control study.

The Auckland Heart Study Validation Project

The 186 Participants for the Auckland Heart Study Validation Project were randomly selected from the control group of the Auckland Heart Study. Of those who could be contacted, 152 completed a seven day physical activity and food intake diary. The seven day diary was the gold standard for the three month physical activity recall questionnaire used in the Auckland Heart Study. The response rate for completing the seven day diary was 82%. The original control group had been randomly chosen from the community and hence the sub-sample of 152 participants represented a reasonable cross-section of the community.

The correlations for the three month recall questionnaire compared with the seven day diary, were 0.61, 0.49 and 0.86 for moderate, vigorous and total activity respectively. These findings were consistent with other validation studies in the literature. One of the strengths of the Auckland Heart Study Validation project was that it was undertaken in the community population for which
it was intended. It was concluded that the three month physical activity recall measured physical activity in general and over the three recall period.

**Auckland Blood Pressure Control study**

Low levels of physical activity have been shown in observational studies to be associated with a high incidence of both coronary heart disease and hypertension. A concern with observational studies is that the findings may be due to confounding factors which are not able to be controlled, either in the design or the analysis. The best method of controlling for confounding is through the use of randomised controlled trials. The literature on physical activity as a means of lowering blood pressure contains many randomised trials and almost all have methodological weaknesses. Moreover, most of those studies have been conducted in laboratory settings; very few trials of physical activity and blood pressure have been conducted in community settings.

The literature on salt restriction as a means of lowering blood pressure contains numerous well designed randomised controlled trials showing that salt restriction can lower blood pressure. While significant results have been achieved from salt restriction most of these studies have been involved intensive input from dietitians. None of the community based studies have demonstrated significant blood pressure reductions.
The aim of the Auckland Blood Pressure Control study was to assess the effectiveness of physical activity and/or salt restriction as therapies to lower blood pressure in treated hypertensive patients in a community setting. The research design was a factorial design randomised controlled trial of physical activity and salt restriction as therapies for lowering blood pressure. Participants were recruited for the study from general practitioners and a variety of public advertisements. The study was conducted over six months and 181 of the baseline 208 participants completed the study. The two interventions were brisk walking for 40 minutes, three times a week and salt restriction advice. The main outcome measures were blinded blood pressures measured at three and six months. The average age of the participants was 55 years and there were approximately equal numbers of men and women. At the three month interview there was a statistically reduction in systolic blood pressure for salt restriction and physical activity as separate therapies, but not for the combination. There was no significant reduction in diastolic blood pressure at the three or six month assessment. Although the Auckland Heart Study three month recall questionnaire was valid for the case-control study there was concern that it was not sensitive enough for the randomised controlled trial. It was concluded that both physical activity and salt restriction lowered systolic blood pressure, at least in the short term, in persons with hypertension treated with medication in a community setting.
Acknowledgements

The studies presented in this thesis were funded by the National Heart Foundation of New Zealand and the Medical Research Council of New Zealand and their support is gratefully acknowledged. The author was also funded for one year of the study by the Maurice and Phyllis Paykell Teaching Scholarship.

Professor Robert Beaglehole was the co-principal investigator for the Auckland Blood Pressure Control Study and I acknowledge his support and involvement. He has also been the supervisor for this thesis and a collaborator on several journal articles and in these tasks he has been a great mentor. I have learnt a considerable amount about research and medical writing as a result of his supervision.

Of special note has been the contribution of Meg Butler who worked with me on the Auckland Heart Study Validation Project and was the project manager for the Auckland Blood Pressure Control Study. In this role she taught me many aspects of the management of research studies. Rochelle Curry joined the Auckland Blood Pressure Control Study once it was in progress and proved to be a most valuable colleague. The author wishes to also thank Nicole Jackson who assisted with questionnaire development and the early part of the blood pressure study.
The author wishes to thank the following persons who helped in the design of the studies: Trevor Beard, Rod Jackson, Stephen MacMahon, Trefor Morgan, Robert Scragg, Norman Sharpe, Olaf Simpson. Susan Sharpe and Sarah Sharpe assisted with the coding of the food intake diaries which turned out to be an enormous task. Thanks also to Greg Gamble who assisted with the programming for the food composition tables, Alistair Stewart who assisted with the statistical analysis and the Dietary Department at Auckland Hospital for the detailed information on salt containing foods.

I wish to thank the participants who willingly took part in the study. I hope the information and advice given to them compensates for the work they have undertaken. Finally I wish to thank Christine and my children, Justine, Nicola and Michael. I have tried not to let this thesis impinge on their lives but inevitably it has.
# List of tables and figures

## Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Summary of major issues considered in this thesis.</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td>Summary of protocols use for validation of exercise questionnaires in chapter 2.</td>
<td>53</td>
</tr>
<tr>
<td>2.2</td>
<td>Gold standards as used in the validation of physical activity measures with the advantages and disadvantages of each method.</td>
<td>55</td>
</tr>
<tr>
<td>4.1</td>
<td>Mean time spent in various activities from 3 month recall by gender.</td>
<td>81</td>
</tr>
<tr>
<td>4.2</td>
<td>Hours per day spent in levels of activity by age and gender.</td>
<td>82</td>
</tr>
<tr>
<td>4.3</td>
<td>Total energy expenditure by age and gender.</td>
<td>83</td>
</tr>
<tr>
<td>4.4</td>
<td>Mean values of moderate, vigorous and total physical activity assessed by 3 month recall and 7 day diary.</td>
<td>84</td>
</tr>
<tr>
<td>4.5</td>
<td>Comparison between 3 month recall and 7 day diary.</td>
<td>85</td>
</tr>
<tr>
<td>4.6</td>
<td>Showing 2x2 table cells for the calculation of sensitivity and specificity and Cohens kappa score.</td>
<td>86</td>
</tr>
<tr>
<td>6.1</td>
<td>Summary of protocols used for blood pressure measurement and physical activity for the 31 reviewed studies.</td>
<td>139</td>
</tr>
<tr>
<td>6.2</td>
<td>Summary of characteristics of physical activity and blood pressure measurement.</td>
<td>142</td>
</tr>
<tr>
<td>6.3</td>
<td>Summary of study design and design bias, resting pre and post-exercise group mean blood pressure and weight</td>
<td>144</td>
</tr>
</tbody>
</table>
6.4 Summary of factors postulated to be part of the mechanism of blood pressure reduction following physical activity

8.1 Baseline mean blood pressures for the initial groups (i.e., includes those who did not complete the study) and sub-group completion rates.

8.2 Values for systolic and diastolic blood pressure, weight, alcohol intake, moderate, vigorous and total activity for all participants at baseline, 3 months and final interview (6 months).

8.3 Baseline Characteristics of participants who completed the study, for age, gender, ethnicity, smoking status, duration of taking medication and alcohol intake.

8.4 Baseline Characteristics of participants who completed the study for alcohol intake.

8.5 Baseline Characteristics of participants in terms of main lifetime occupation, current working situation and education.

8.6 Baseline Characteristics of participants in terms of marital status and source of recruitment in those who completed the study.

8.7 Baseline medication characteristics of participants who completed the study.

8.7a Baseline numbers of medication by intervention group. The denominator is medication type.

8.8 Baseline characteristics for the participants who completed the study.

8.9 Characteristics at end of study for the participants who completed the study.
8.10 Values for systolic and diastolic blood pressure, weight, alcohol intake, moderate, vigorous and total activity 24 hour urine sodium, salt frequency score for groups A&B compared with groups C&D.

8.11 Differences and difference of the differences from the baseline to 3 month; for systolic blood pressure for the exercising groups (A+B) compared with the non exercising groups (C+D).

8.12 Systolic blood pressure at baseline, 3 months and final interview for exercising versus non-exercising groups.

8.13 Differences and differences of differences from baseline to 3 month diastolic blood pressure for the exercise groups (A+B) compared with the non-exercising groups (C+D).

8.14 Diastolic blood pressure at baseline, 3 months and final interview for exercising versus non-exercising groups.

8.15 Values for systolic and diastolic blood pressure, weight, alcohol intake, moderate, vigorous and total activity, 24 hour urine sodium, salt frequency score: for groups A&C compared with groups B&D; at the baseline, 3 months and final interview;

8.16 Differences and differences of differences from the baseline to the 3 month systolic blood pressure for the salt restriction groups (A+C) compared with the non salt restricting groups (B+D).

8.17 Systolic blood pressure at baseline, 3 months and final interview for the two salt restriction groups versus the two non-salt restricting groups.
8.18 Differences and differences of the differences from baseline to 3 month diastolic blood pressure for the salt restriction groups (A+C) compared with the non-salt restricting groups (B+D).

8.19 Diastolic blood pressure at baseline, 3 months and final interview (6 months); for salt restricting groups (A+C) versus non-salt restricting group.

8.20 Differences and differences of differences from baseline to 3 month systolic blood pressure for the intervention groups (A,B,C) compared with the control group (D).

8.21 Systolic blood pressure at baseline, 3 months and final interview (6 months).

8.21a Mean of the differences in systolic blood pressure at baseline and 3 months, by medication subgroup

8.22 Difference between systolic blood pressure at baseline, 3 months and final interview by sub-groups.

8.23 Distribution of change in systolic blood pressure at 3 months by alcohol intake for intervention sub-groups versus control group.

8.24 Differences and differences of differences from baseline to 3 month diastolic blood pressure for the intervention groups (A,B,C) compared with the control group (D).

8.25 Diastolic blood pressure at baseline, 3 months and final interview.

8.26 Difference between diastolic blood pressure at baseline, 3 months and final interview.
8.27 Numbers of participants reporting changes in systolic blood pressure at baseline, 3 month and 6 month interview.

8.28 Numbers of participants reporting changes in diastolic blood pressure at baseline, 3 months and final interview.

8.29 Mean numbers of days between baseline and 3 month; 3 month and six month interviews.

8.30 Physical activity values at baseline, 3 months and final interview.

8.31 Hours since last physical activity by group; at baseline, 3 month and six month interview.

8.32 Alcohol, weight and pulse by group, at baseline, 3 months and final interview.

8.33 Twenty four hour sodium excretion and salt frequency results; baseline, 3 months and final interview; comparison of salt restriction groups versus no salt restriction.

8.34 Twenty four hour sodium excretion and salt frequency results; baseline, 3 months and final interview; comparison of salt restriction groups versus no salt restriction at 6 months.

8.35a Total cholesterol, triglyceride, low density and high density lipoprotein baseline and final interview.

8.35b Total cholesterol, triglyceride, low density and high density lipoprotein baseline and final interview.

8.36 Characteristics at 6 months, by age, for the participants who completed the study.
8.37 Blood pressure at baseline, 3 months and final interview by age and intervention group.

8.38 Differences and differences of differences for systolic blood pressure from baseline to 3 month and six months; for the participants aged 55 years or less compared with the participants aged over 55 years.

8.39 Diastolic Blood pressure at baseline, 3 months and final interview by age and intervention group.

8.40 Differences and differences of differences from baseline to 3 months and six months for diastolic blood pressure for the participants aged 55 years or less compared with the participants aged over 55 years.

8.41 Baseline characteristics, by gender, for the participants who completed the study.

8.42 Characteristics at 6 months, by gender, for the participants who completed the study.

8.43 Differences and differences of differences from baseline to 3 months and six months; for systolic blood pressure for women compared with men.

8.44 Differences and differences of differences from baseline to 3 months and six months for diastolic blood pressure for the women compared with men.

8.45 Blood pressure at baseline, by gender, 3 months and final interview.

8.46 Systolic blood pressure at baseline, 3 months and final interview (6 months) by gender and intervention group.)
8.47 Diastolic Blood pressure at baseline, 3 months and final interview by gender and intervention group.

8.48 Changes in medication at three months, six months and overall changes.

8.49 Prescribed medication at baseline, 3 months and 6 months converted to units of medication score.

8.50. Prescribed medication converted to units of medication score for difference from baseline to 3 months and 6 months.

8.51 Non-pharmacological therapies tried by the participants before the study

8.52. Summary of the differences between group A and groups B and C.

**Figures**

1.1 The chronological order for the thesis topics. 4

1.2 Types of epidemiological studies in descending order of the strength of controlling for confounding 12

1.3 Time sequence for a randomised controlled trial 14

3.1 Overview of Auckland Heart Study Validation Project 59

3.2 Time sequence of Auckland Heart Study Validation Project 64

7.1 Overview of the ABC Study 153

7.2 Factorial design 154
8.1 Mean values for systolic and diastolic blood pressure for all participants at baseline, three months and six months.

8.2 Systolic blood pressure at baseline, 3 months and final interview for two exercise groups versus the two non-exercising groups.

8.3 Diastolic blood pressure at baseline, 3 months and final interview for two exercise groups versus the two non-exercising groups.

8.4 Systolic blood pressure at baseline, 3 months and final interview for the two salt restriction groups versus the two non-salt restricting groups.

8.5 Diastolic blood pressure at baseline, 3 months and final interview; for salt restriction versus no salt restriction.

8.6 Systolic blood pressure at baseline, 3 months and 6 months for the four study sub-groups: group A, exercise and salt restriction; group B exercise only; group C salt restriction only; group D control group.

8.7 Diastolic blood pressure at baseline, 3 months and 6 months for the four study sub-groups: group A, exercise and salt restriction; group B exercise only; group C salt restriction only; group D control group.

8.8 Moderate activity in kcal/kg/day at baseline, 3 months, 6 months by the four intervention subgroups.

8.9 Total activity in kcal/kg/day at baseline, 3 months, 6 months by the four intervention subgroups.

8.10 Mean systolic blood pressure in mm Hg by participants aged 55 years and younger versus those aged over 55 years.
8.11 Mean diastolic blood pressure in mm Hg by participants aged 55 years and younger versus those aged over 55 years.

8.12 Systolic blood pressure (in mm Hg) at baseline by gender at baseline, 3 months and six month interview.

8.13 Diastolic blood pressure at baseline by gender at baseline, 3 months and six month interview.
Chapter 1: Introduction

1.1.1. General Introduction
1.1.2. Hypertension, salt intake, physical activity and coronary heart disease
1.1.3. Study overview
1.1.4. Aims: Auckland Heart Study Validation Project
1.1.5 Aims: Auckland Blood Pressure Control Study
1.2.1. Basic principles in a randomised controlled trial
1.2.2. General features of a randomised controlled trial
1.2.3. Ethical issues in a randomised controlled trial
1.2.4. Validity
Chapter 1: Introduction

1.1.1. General Introduction

The historical background to this thesis starts with the MONICA project; a 10 year international study of coronary heart disease and associated risk factors. As part of the MONICA project a case control study, known as the Auckland Heart Study, examined risk factors and their relationship to coronary heart disease. The Auckland Heart Study was conducted in Auckland from 1986-1988. The results from that study confirmed that leisure time physical activity was associated with lower rates of coronary heart disease. Having found this result the Auckland Heart Study investigators wished to validate the physical activity questionnaire. The validation process was known as the Auckland Heart Study Validation Project and the physical activity validation is described in chapters 2 to 5 of this thesis.

The validated questionnaire was subsequently used in the Auckland Blood Pressure Control Study. This was a factorial design, randomised controlled trial of physical activity and salt restriction as an adjunct to pharmacological treatment of hypertension in persons treated in a community setting. Although the validation study is presented at the beginning of this thesis the major focus is on the Auckland Blood Pressure Control Study which is presented in chapters 6 to 9. Physical activity and its role in
lowering blood pressure is a major theme of this thesis while salt restriction and blood pressure reduction receives less attention. A summary of the major issues in the study are shown on table 1.1 while the chronological order is shown in figure 1.1.

Table 1.1 Summary of major issues considered in this thesis

-(in order of importance)

(i) Exercise in the management of hypertension- Auckland Blood Pressure Control Study.

(ii) Validation of an exercise questionnaire- Auckland Heart Study Validation Project.

(iii) Salt restriction as management for hypertension- Auckland Blood Pressure Control Study.

(iv) Validation of a food intake questionnaire- Auckland Heart Study Validation Project.
Fig 1.1: The chronological order for the thesis topics.

Auckland Heart Study 1986-88

Validation of an exercise questionnaire - this project included a food intake validation

Auckland Heart Study Validation Project 1988-89

Auckland Blood Pressure Control Study 1989-90

Randomised controlled trial of exercise and salt restriction in treated hypertensives
1.1.2. Hypertension, salt intake, physical activity and coronary heart disease

This section overviews the relationship of physical activity and salt restriction to hypertension as well as the relationship of hypertension to coronary heart disease. Full details of the relevant literature are presented in chapter 2 and 6. Hypertension is a major risk factor for coronary heart and cerebrovascular disease (Fraser, 1986). In New Zealand hypertension is the most common reason for visits to a general practitioner for the 35-64 age group (Squires, 1983), and in 1986 the cost of prescription medication was $80 million out of the $506 million total pharmaceutical bill (Sinclair, 1989). Preliminary results from the 1991-1992 Waimedca study, which obtained data from a cross-section of general practitioners in Hamilton, New Zealand showed that hypertension was the commonest diagnosis out of forty diagnostic categories. Hypertension was also the second most common reason for a visit to the practice nurse after immunization and there were six antihypertensive medications in the top fifty prescribed items (McAvoy, 1992). The Life in New Zealand Study examined a random sample of 3015 New Zealanders over the age of 15 using the World Health Organization criteria for hypertension {i.e. systolic pressure equal to or greater than 160 mm Hg and a diastolic pressure equal to or greater than 95 mm Hg} (Nye, 1992). The study found that 29% of men and 24% of women over the age of 45 years were hypertensive by this criteria. Regardless of blood pressure 24% of men and 33% of women over the age of 45 years were on drug treatment.
for hypertension. For all age groups 9% of men and 12% of women were on treatment for hypertension.

Physical activity has been shown in observational studies to be associated with reduced risk of developing hypertension and coronary heart disease (LaPorte, 1984); (Paffenbarger, 1978); (Paffenbarger, 1983). Physical activity is regarded as a potential adjunct and an alternative to pharmacological therapy yet its efficacy in treating hypertension has usually been shown in laboratory settings (Arroll, 1992,); (Kaplan, 1985); (McMahon, 1985). Few studies have included participants on medication and most have been studied in laboratory settings. The majority of studies with exercise as a therapy for hypertension have important methodological flaws.

In spite of a large amount of research, the issue of salt intake and its relationship to hypertension is still controversial. Salt restriction as a treatment for hypertension has been demonstrated in randomised trials (MacGregor, 1982); (MacGregor, 1989) and much of the controversy comes from observational studies in which the relationship between salt intake and hypertension has been weak (Smith, 1988); (Elliot, 1989). Recent reviews of the salt and hypertension literature may resolve the controversy as they found consistent high quality evidence in favour of salt as a causative agent in hypertension (Law -III, 1991); (Law-part II, 1991). Salt reduction is an important non-pharmacological therapy for individuals with hypertension; a
population approach is also possible primarily by a reduction in salt content of manufactured foods.

Coronary heart disease is the leading cause of death in New Zealand being responsible for 26.0% of male and 24% of female deaths in 1988 while cerebrovascular disease accounted for 7.0% of male deaths and 13% of female deaths (National Health Statistics Centre, 1988). New Zealand's high prevalence of coronary heart disease is most likely due to factors including a high intake of dietary saturated fat, smoking, hypertension and lack of regular physical exercise all of which can be modified (National Heart Foundation of New Zealand, 1983). Pharmacological therapy for hypertension has been less successful in preventing coronary heart disease than cerebrovascular disease (Collins, 1990). In a recent meta-analysis of pharmacological trials for treating hypertension the reduction in coronary heart disease was found to be 14% whereas 20-25% would have been expected from observational epidemiological studies (Collins, 1990); (MacMahon, 1990). It is not certain why there is this discrepancy although one possibility is the elevation of serum cholesterol induced by some blood pressure medications. As physical activity can lower blood pressure and/or serum cholesterol there is the potential for greater reductions in coronary heart disease than obtained with medication, at lower cost, and without the possible risks of pharmacological side effects (Arroll, 1992; Wood, 1990).
Justification for the Auckland Heart Study Validation Project and the Auckland Blood Pressure Control Study:

Observational data has shown that there exists an inverse relationship between physical activity and hypertension/coronary heart disease. The validation study was necessary to assess the validity of the Auckland Heart Study results. There is always a concern with observational data that the findings may be the result of confounding with other variables. To clarify some of these issues randomised controlled trials are necessary. The justification for the Auckland Blood Pressure Control Study is that studies showing that exercise can reduce blood pressure have been performed only in laboratory or supervised settings; it was not known if such interventions could be successful in a community setting. Relatively unsupervised salt intervention studies have been attempted in community settings but have not significantly reduced blood pressure.

In summary, hypertension is a significant contributor to morbidity, mortality and economic cost of health care in New Zealand. Validation of the physical activity questionnaire from the Auckland Heart Study was necessary to confirm that it was truly measuring physical activity. Regular physical activity and salt restriction are important areas of research in the study of hypertension as both have the potential to lower levels of blood pressure in the community. The Auckland Blood Pressure Control
study examined the effectiveness of unsupervised exercise and salt restriction interventions on blood pressure in community settings.

1.1.3. Study overview

Auckland Heart Study Validation Project (1988)
The Auckland Heart Study Validation Project involved the validation of the 3 month physical activity recall and a 3 month food frequency questionnaire used in the Auckland Heart Study. The validation of the food frequency questionnaire was important for the Auckland Heart Study but only a minor part of the validation study. In this context the food intake data was an additional source of concurrent validation for the exercise questionnaire.

The Auckland Blood Pressure Control study (1989-1990)
The Auckland Blood Pressure Control study was a factorial design, randomised controlled trial of physical activity and salt restriction as an adjunct to pharmacological treatment of hypertension in persons treated in a community setting. It was conducted in Auckland between September 1989 and July 1990. The questionnaire used to measure exercise in the study was that used in the Auckland Heart Study, a case-control study of coronary heart disease.
1.1.4. Aims: Auckland Heart Study Validation Project
The aims of the Auckland Heart Study Validation Project were:
(1) To validate the physical activity section of the Auckland Heart Study using a physical activity and food intake diary as the gold standard. As well as direct validation it was possible to have an additional validation between the (diary) food intake and physical activity (from the 3 month recall and 7 day diary) output (i.e. calories consumed as food versus calories expended as physical activity).

1.1.5 Aims: Auckland Blood Pressure Control Study
The aim of the Auckland Blood Pressure Control Study was to investigate the effect of regular moderate exercise and/or salt restriction on blood pressure levels in a community based sample of persons with treated hypertension. The Auckland Blood Pressure Control Study assessed:

(i) The effect of regular moderate exercise (brisk walking) and/or salt restriction on blood pressure levels in persons with treated hypertension.
(ii) Other changes resulting from an exercise programme, i.e., weight, pulse, diet, alcohol use and serum cholesterol.
(iii) The compliance with an exercise and/or a salt restriction programme for a group of persons with treated hypertension.
1.2.1. Basic principles in a randomised controlled trial

This section describes the methodological issues relevant to randomised controlled trials as the major focus of this thesis is the Auckland Blood Pressure Control Study. A major task of epidemiology is to evaluate the causes of disease and evaluate the benefits of interventions. The most powerful method of establishing cause and effect is the randomised controlled trial (Department of Clinical Epidemiology and Biostatistics. McMaster University part IV, 1981) (fig 1.2). The main strength of randomised trials in the context of exercise and blood pressure is to reduce the potential for confounding by randomising subjects to intervention and control groups. Confounding in the context of exercise and blood pressure could result from healthy people having lower blood pressure and also being more likely to want to participate in exercise. People who exercise are also more likely to eat healthier foods and smoke less than non-exercisers and these factors are also potential confounders. It is therefore possible that healthy people have lower blood pressure not because of their physical activity but due to other lifestyle and genetic factors. If the association of lower blood pressures with more active persons was only due to confounding in observational studies then exercise would not lower blood pressure in randomised trials. Observational studies attempt to control for confounding by adjustment in the analysis although it is not
possible to adjust for confounding completely. Randomisation is the best method of controlling for confounding.

Fig 1.2 Types of epidemiological studies in decreasing order of the strength of evidence on causation

1. Randomised controlled trial  
   e.g Auckland Blood pressure control study
   
   experimental ↑
   
   observational ↓

2. Cohort study  
   e.g. Harvard Alumni Study

3. Case-control study  
   e.g. Auckland Heart Study

4. Cross sectional study  
   e.g. Hillary Commission  
   Life in New Zealand Study

5. Ecological study  
   e.g Seven countries  
   study by Keys et al

6. Case series
1.2.2. General features of a randomised controlled trial:

A randomised controlled trial (also known as a clinical trial) involves a group of persons who have been selected by uniformly applied inclusion and exclusion criteria. This group is then randomised into treatment/exposure sub-groups and it is the investigator who controls the allocation of the treatment, not the participants (fig 1.3). In observational studies participants can choose their intervention (i.e. being physically active) whereas this is randomly allocated by the investigator in a randomised trial. This is the major difference between experimental and observational studies.

There are six pre-requisites for a successful randomised controlled trial (Sackett, 1983). These are:

(i) There is a legitimate need for the trial.
(ii) The question posed is both appropriate and unambiguous.
(iii) The trial architecture is valid.
(iv) The need to strike a balance between efficiency and generalizability.
(v) The trial protocol is feasible.
(vi) The trial administration is effective.
Figure 1.3: Time Sequence for a randomised controlled trial

Key

R = randomisation

exposed = treatment group (for the ABC study: exercise and salt restriction)

not exposed = control or placebo group

outcome = (for ABC study) level of blood pressure
(i) The trial needs to be done.

This prerequisite has 3 parts. First, the clinical disorder should be an important condition in that it has a significant mortality and morbidity. This has been outlined for hypertension in section 1.1.2. Second, the existing intervention should be of limited usefulness in terms of efficacy, safety and cost. The current mainstay of treatment for hypertension is pharmacological with a minority of investigators suggesting physical activity as an alternative or adjunct to medication. There is no New Zealand study which examines the frequency with which doctors prescribe physical activity as a treatment for hypertension. Anecdotal evidence suggests that physical activity is not currently routinely advised as an adjunct to pharmacological treatment and hence it is reasonable to investigate this aspect. Also there have been no major studies of exercise as an adjunct to pharmacological treatment in the community setting and therefore it is a worthwhile subject for investigation. Third, a new intervention should have a sufficiently high probability of benefit. As physical activity does more good than harm in terms of coronary heart disease (Arroll, 1989) and exercise has been shown in laboratory settings to lower blood pressure it is reasonable to investigate its application in the community.
(ii) The question posed is both appropriate and unambiguous. The research question for the Auckland Blood Pressure Control Study was as follows: Is regular moderate exercise and/or salt restriction effective in further lowering blood pressure in a community based sample of persons with drug treated hypertension? This is considered to be an appropriate and unambiguous research question.

(iii) The trial architecture is valid.

Internal validity is a major issue in the validity of the trial architecture and the chief device for achieving this is random allocation. In a randomised controlled trial it is the investigator who controls the allocation of treatment and not the participants. Random allocation is the key distinction between an experimental and observational study and eliminates confounding that may occur in observational studies. Bias is defined as systematic deviation from the truth and a major source of bias in observational studies is known as confounding. This is where healthier persons are more likely to participate in healthier activities thereby falsely associating that activity with good outcomes and vice versa (see 1.2.1). The prospective nature of randomised controlled trials removes the problem of recall bias which occurs where persons with serious illnesses are more likely to remember exposures than healthy controls. Randomisation has other advantages in that, as sample size increases, the unknown confounding variables in a trial are likely to be balanced between groups. Randomisation is also of
value in analysis as many statistical tests rest upon the assumption of random allocation.

Unblinding or unmasking is where the assessors or the participants in a study become aware of the intervention and consequently alter their behaviour. Lack of blinding is a form of bias that occurs in observational studies and can occur in a poorly designed randomised controlled trial. A way of circumventing this is to use the method of double blinding and trials involving medication lend themselves to this approach. This is achieved by having both the participant and the assessors unaware of the treatment being given. In interventions such as physical activity, which cannot be blinded, the outcome measurement needs to be performed in a blind fashion. For an outcome such as blood pressure this requires an assessor, blind to the intervention, to measure the blood pressure. Blinding can be achieved in studies of salt restriction by ensuring that all participants eat a low salt diet and randomly allocating salt and placebo containing tablets (MacGregor, 1989).

Randomised controlled trials are subject to contamination and co-intervention. Contamination occurs when the control group receives some or all of the intervention. This is a major issue with unblinded or behavioural change studies where it is possible for the control group to guess or know the intervention and, unbeknown to the investigators, commence the intervention activity. Contamination results in an under-estimate of the effect between intervention and control groups. Co-intervention occurs when the
intervention group obtains an additional intervention such as extra contact with the investigators. A common example of co-intervention is where the intervention group is treated as inpatients while the control group is seen in the outpatient department. This form of bias can result in an over-estimation of the effect of the intervention. Examples of co-intervention in the physical activity literature are described in chapters 6 and 9.

Although it seems that randomised controlled trials have many advantages they also have disadvantages such as: being expensive in terms of time and money; participants may not be representative of all patients; and a potentially effective treatment is withheld from some subjects while others may be exposed to a possibly dangerous treatment (Streiner, 1989).

(iv) The inclusion/exclusion criteria: to strike a balance between efficiency and generalizability

All studies involve a balance between efficiency and generalizability. Generalizability is also known as external validity. To gain maximum efficiency the smallest number of participants are required to give a result. This would require participants to be people at highest risk of an event or, in studies of hypertension, those with the highest initial blood pressure. It is also desirable to study people most likely to respond to therapy and in studies of behavioural change would include persons likely to comply with the intervention. The concern with treating a high risk/high response group is that they will be a minority of those
with the condition of interest and the trial results may not be generalizable. Limitations to generalizability are inevitable and one way of estimating this is to document those who are excluded, withdraw or cross-over. The term cross-over refers to those participants who change from an intervention group to a control group or vice-versa.

(v) The trial protocol is feasible

There are four parts to this prerequisite. First, the study must be seen to be necessary for potential clinical collaborators (and potential participants) to contribute patients and their time. Second, the intervention should be able to be implemented in everyday practice. The third issue is that of the demands on the potential clinical collaborators. They may only be required to provide patients for the study or at the other extreme to require them to complete extensive documentation and perform frequent follow up visits. The collaborators must be able to see some personal reward for the effort invested. Fourth, the availability of adequate numbers of patients for the study is necessary. This can often be more difficult than anticipated and sample size estimations may need to be revised in terms of the acceptable type 1 and type 2 statistical errors. Thus the study needs to be attractive to the participants as well as their medical practitioners. When recruitment is a problem solutions can be found by using a paired and/or cross-over design. When all this fails a multi-centre design, with its additional administrative
burden and expense, can be considered. An initial pilot study can give a preview of recruiting difficulties but this may raise other difficulties such as the ethics of studying persons whose results may not be able to be used in final analyses. Pilot studies can also "consume" participants and if they are in short supply can scuttle the definitive study. The use of participants for pilot studies can have a deleterious effect on recruitment when there is a shortage of potential candidates.

(vi) The trial administration is effective

The effective administration of a study is necessary for the completion of the study and to maintain its validity and credibility. Administrative problems can be more of a problem than scientific ones. The study protocol must be fully documented in a study manual that contains unambiguous rules for handling the many issues that arise during a study. Lines of responsibility and authority must be clearly specified. Publication policies and rules for terminating the study need to be established at the outset.
1.2.3. Ethical issues in a randomised controlled trial

General Ethical Issues

Ethical issues are important in all research, especially when experimental interventions and control groups are involved. Randomised controlled trials have specific issues as the investigator is withholding a potentially beneficial/harmful treatment from one group in a study.

There are three basic ethical principles that need to be considered (Lebacqz, 1983). These are:
(i) Respect for persons.
(ii) Beneficence.
(iii) Fair distribution of benefits and burdens.

According to Lebacqz (1983) the four part Nuremberg code serves as a baseline for assessing the ingredients for giving valid consent to participate in a clinical trial. These are that a person must:
(i) Have legal capacity to consent.
(ii) Be sufficiently free so that consent is truly voluntary.
(iii) Be given adequate information on which to decide.
(iv) Understand that information well enough to make an enlightened decision.
Current ethical consideration would emphasize the issues of informed consent and the right to withdraw without any prejudice to future treatment.

Specific issues for randomised controlled trials

Experimental research can be justified on the grounds that it can produce benefits and prevent harmful practices. The risk-benefit analysis is an important part of justifying a study and risks are only acceptable if they are outweighed by potential benefits. It is also important to separate research from treatment so as to not confuse the two processes. Equally, it is important not to underestimate the risks of standard treatment as this too contains many risks.

Randomisation into treatment groups presents special problems for randomised controlled trials. First, should potential participants be informed that their treatment will be randomly allocated? In most studies this presents no problem. However when participants have a perception that one treatment is superior to another problems can arise. One solution is to pre-randomise. This involves participants being randomised into a treatment group that will receive standard treatment and a consent group that will be asked to consent to a new treatment (Zelen, 1979). The standard treatment group is the control group but does not necessarily know that they are in a particular study. They may be told that they are in a research study but would not be aware that persons in other arms of the study are being offered different interventions. This
approach was necessary in a study of total mastectomy versus segmental mastectomy in order to recruit enough patients (Fisher, 1985). Understandably, very few women would choose to have a total mastectomy if there was any doubt as to its efficacy. If the new therapy could be shown to be convincingly better than or as good as total mastectomy as result of a well conducted randomised trial there would be grounds for recommending the less destructive therapy. Without the pre-randomisation it would not have been possible to perform the trial and the benefits of segmental versus total mastectomy could not subsequently be offered to women with breast cancer.

The issue of having a control group raises its own ethical dilemmas. Is it right to withhold a known treatment in order to expose some persons to an unvalidated treatment? Conversely is it right to withhold a new therapy that has fewer side effects while maintaining some of the study participants on the known therapy? The use of randomised controlled trials and control groups can be justified in terms of gaining better quality of information than is possible with other methods of investigation. In general, the new therapy would be expected to have at least as good risk-benefit ratio as the standard therapy. As a safeguard it is becoming standard practice to have an ethical evaluation of the results during a study. Criteria can be set in advance and if an imbalance of benefit/harm appears the study can be terminated. This does not necessarily quash all debate as the Physicians Health Study, investigating the primary prevention of myocardial infarction
using aspirin, was terminated early and considerable debate ensued (The Steering Committee of the Physicians Health Study Research Group, 1989). Although the published debate favoured the aspirin intervention there was concern that assessment of the cerebrovascular adverse effects was limited by early termination (Fuster, 1989).

Ethical approval

Ethical approval for the Auckland Heart Study and the Auckland Blood Pressure Control study was obtained from the University of Auckland Human Subjects Research Ethical Committee.

1.2.4. Validity

There is a concern in all research that a measure may not be measuring what it is thought to be measuring. This is known as validity. There are five categories of validity: repeatability; face validity; content; criterion; and construct validity (Streiner, 1989).

(i) Repeatability: Repeatability is a measure of the extent to which a measure is reproducible. The reproducibility can be assessed in different situations or with different observers. A measure which is not reproducible cannot be valid, although a valid measure must be reproducible. In other words repeatability is a necessary but not sufficient condition for validity. In the Auckland
Heart Study Validation Project repeatability was not considered as this was done as part of the main Auckland Heart Study.

(ii) Face validity: Face validity is where a measure looks like a reasonable measure and is considered the weakest form of validity. Its credibility can be enhanced when there is agreement amongst a panel of experts. To assist this process for the Auckland Heart Study Validation Project the exercise diary was shown to other researchers in the exercise field for comment. A literature review (chapter 2) and checking other questionnaires also assisted in ensuring that face validity was achieved.

(iii) Content validity: Content validity is where a measure appears to contain all the important elements. Achieving content validity for the exercise diary involved a similar process to that for establishing face validity. In order to ensure that all activities were included activities were subdivided into leisure, paid work and home activities. The diary also consisted of lists of commonly performed activities as a memory aid to participants (appendix A). The use of the Five City format, (see chapter 3, methods) where time spent in light intensity activities is what remains after accounting for sleep, moderate and hard activities, can be criticized in that it does not seem to account for the details of light activity. Other levels of activity, such as tremor, are beyond the ability of measurement by questionnaire.
(iv) Criterion validity: This is the strongest approach to validity as it involves the assessment of a measure against a gold standard. It is divided into two forms which differ in time. Concurrent validity is where the comparison is made at the same time and predictive validity is where a measure is able to predict future status. The Auckland Heart Study Validation Project used the seven day diary as the gold standard for physical activity and hence it was a measure of concurrent validity. The comparison of the total physical activity output with the total calories from the food consumed provided an additional measure of concurrent validity. There was no attempt to assess the predictive validity of the exercise questionnaire although some of the questionnaires on which it was based have predicted disease outcomes (Paffenbarger, 1986).

(v) Construct validity: This is the most difficult to understand of the forms of validity. It is used in circumstances where there is no other measure of the attribute under study. A theoretical construct between the measure of interest and other measures is made and then compared to see if the results have similar values and trends. If they do not there is no way of knowing whether the measure or the theory is wrong. As the seven day diary is a dubious gold standard (see chapter 2 literature review and chapter 3 methods) it could be argued that the diary was a measure of construct validity rather than concurrent validity. For the Auckland Heart Study Validation Project this was doubly difficult as there
is little expert agreement on the best gold standard for food and physical activity.
Chapter 2: Literature review: Auckland Heart Study Validation Project

2.1.1 Introduction
2.1.2 Overview of physical activity literature
2.1.3 Validation of physical activity studies
2.1.4 Seven day recall and seven day diaries
2.1.5 Studies with fitness as a gold standard
2.1.6 Studies of motion activity sensors
2.1.7 Validation of "simple questions"
2.1.8 A study of multiple methods of measuring physical activity
2.1.9 Repeatability
2.1.10 Conclusions
Chapter 2: Literature review: Auckland Heart Study Validation Project

2.1.1. Introduction
This chapter describes the literature review for the Auckland Heart Study Validation Project. As stated in chapter 1 the Auckland Heart Study Validation Project was a study designed to validate the 3 month physical activity recall questionnaire using a seven day prospective diary. The Auckland Heart Study physical activity questionnaire was subsequently used as the measure of physical activity in the Auckland Blood Pressure Control Study. This review deals with some of the general issues involved in the validation of physical activity questionnaires as well as reviewing other studies of physical activity validation. (Tables 2.1 and 2.2 are included at the end of the chapter: p 53 to p 56)

2.1.2. Overview of physical activity literature
The role of physical activity in the primary prevention of coronary heart disease has been the subject of numerous epidemiological studies. Most studies conclude that physically active individuals are at lower risk for coronary heart disease than less active persons (Powell, 1987). A critical review of the literature using pre-determined criteria for the causal relationship between
physical activity and coronary heart disease concluded that physical activity is inversely and causally related to the incidence of coronary heart disease (Powell, 1987). The range of relative risks for coronary heart disease and inactivity was from 1.5 to 2.4. A meta-analysis of the studies in the Powell et al (1987) review showed higher relative risks between physical inactivity and coronary heart disease for the better designed studies (Berlin, 1990).

Having determined that measuring physical activity in population studies of coronary heart disease is important the next question is how best to measure such behaviours. More than 30 different methods (LaPorte, 1985) have been used to assess physical activity and these fall into seven major categories: calorimetry, job classification, survey procedures, physiological markers, behavioural observation, mechanical and electronic monitors, and food intake measures. LaPorte et al (1985) concluded that: "No single instrument fulfils the criteria of being valid, reliable, and practical while not affecting behaviour." The instruments that are very precise take a considerable time to administer and are therefore impractical for large scale studies. Short surveys are the most practical approach in large-scale studies because they are relatively cheap and simple to administer. However little is known about their reliability and validity over long periods of time. Despite the difficulty of measurement, relatively strong associations have been found between self reported physical activity and health suggesting that, with
improvements in assessment, even stronger associations could be seen. A stronger association between coronary heart disease and inactivity would probably be found if a broad range of activity could be studied. In industrialized countries this is difficult as most people are relatively inactive (LaPorte, 1984). One solution could be to include subjects at the extreme levels of inactivity e.g. quadriplegics and paraplegics.

There are two basic models of association between activity and heart disease (LaPorte, 1984). Firstly, the direct effect with physical activity producing cardiac fitness (i.e., improved collateral circulation) resulting in a decreased risk of heart attack. The other model is that of an indirect effect in which physical activity produces risk factor changes (i.e. lipid and blood pressure changes) thereby decreasing the risk of myocardial infarction. LaPorte et al (1984) favour the latter as the level of activity necessary to achieve cardiovascular fitness is greater than that required to increase high density lipoprotein cholesterol. Others maintain that it is exercise that improves fitness and that aerobic exercise is apparently protective with the proviso that such vigorous activity is defined in individual terms (Smith, 1992). This means that brisk walking cannot always be defined as moderate activity; for a young marathon runner it would be light activity yet vigorous for a 90 year old person (Arroll, 1991a). The epidemiological evidence favouring vigorous activity as being the most beneficial for coronary heart disease may be due in part to bias as vigorous activity is easier to measure than activity of lower intensity. The
large intra and inter-individual variations in physical activity favour such a finding when in reality there may be either a gradient of benefit with increasing physical activity or a threshold effect. In a review of physical activity as a means of lowering blood pressure vigorous activity (70% maximal oxygen uptake max) had no advantage over lower levels of physical activity (50% maximal oxygen uptake max) although most of the studies had sub-optimal study designs (Arroll, 1992). This information has implications for the validation of a physical activity questionnaire in that assessment of physical fitness by a maximal physical activity gold standard may not be appropriate. It is still not clear what kinds and pattern of physical activity are beneficial for the long term prevention of coronary heart disease. 

Although the majority of persons in the community obtain their moderate or vigorous physical activity through leisure activities it would be contrary to the requirements of content validity to exclude work related activity. Authors, such as Taylor et al (1978), have limited their interest to leisure time activities and not unexpectedly have been criticised for such a limitation. From the point of view of the predictive component of criterion validity not all physical activity questionnaires have been predictive of coronary heart disease. In some cases the studies have not be long enough to examine mortality end points. In the case of the Five-City project there has been insufficient time for a significant number of outcomes to have occurred (Sallis, 1985). The Five-City project is an intervention study to modify risk factors for health in general and coronary heart disease in particular and has been conducted in the five Californian communities of Salinas, Monterey, Modesto and San Luis Obispo. A random sample of 2504 persons aged 11-74 have been interview yielding an acceptance rate of 65.73%. The questionnaire format developed by Sallis et al (1985) in the Five-City project seems to
be favoured as the technique of the future by a number of reviewers (LaPorte, 1985); (Washburn, 1986) due in part to its simplicity and brevity. This is achieved by assuming that time spent in low level activity is the remaining time in a 24 hour day, after accounting for time in sleep, moderate and vigorous activities. Low level activity is typically the longest of the activity categories.

2.1.3. Validation of physical activity studies
This section will discuss studies that attempt to validate physical activity questionnaires using the concurrent measure of criterion validity. Studies that use the predictive component of criterion validity, such as those in the Powell et al (1987) review are not considered with the exception of the study by Mundal et al (1987) as this study also used two other measures of concurrent validity. This section is therefore limited to studies examining the concurrent component of criterion validity. Each article is critically appraised. Although some studies are more relevant than others it is not possible to divide the studies into high and low quality in terms of research design as has been possible for randomised controlled trials in chapter 6. This is due to the lack of a gold standard for physical activity and the absence of any universally agreed upon approach to validation. A summary of the appraisal of all the studies is discussed in table 2.1. Table 2.2 contains a summary of the gold standards used in this review with their respective advantages and disadvantages.
2.1.4. Seven day recall and seven day diaries

In a study which attempted to validate three measures of physical activity, the Caltrac activity monitor, the Stanford Physical activity recall and a seven day prospective diary of physical activity were compared (Williams, 1989). The Caltrac is an electronic accelerometer to detect motion. It is 3 x 5 x 0.5 inches in size and has a digital numeric display. It has five pressure sensitive keys which permit the investigator to choose a variety of functions (e.g. resting metabolic rate). This study included 21 men and 24 women with an average age of 24.7 years who were recruited from graduate classes in psychology. The participants received course credits for being involved in this study and the prospective daily diary was regarded as the gold standard. The correlation coefficient for the monitors with the diary was 0.11 in the first week, 0.37 and 0.34 for the second and third weeks. The correlation coefficients for the Stanford questionnaire with the 7 day diary was 0.68, 0.84, and 0.82 for the first, second and third weeks respectively. This improvement over time in the correlation figures may mean that the students became more skilled at filling in the two questionnaires. The well educated nature of the participants and the highly supervised nature of the study make the results difficult to generalize to the wider community. The authors of this paper claimed a high reliability and validity for the Stanford questionnaire. They suggested that Caltrac monitors not be used for the monitoring of long term studies of physical activity because of their high cost, repeated mechanical failures and poor reliability and validity.

Observed behaviour is a potential gold standard for physical activity questionnaires although the process of observation is very labour intensive and may influence the participants and these are major disadvantages. This technique was used in a study to validate a seven day recall questionnaire of physical activity and included 11 overweight boys at a summer camp (Wallace, 1985). Their physical activity was observed by camp counsellors every 15
labour intensive and may influence the participants and these are major disadvantages. This technique was used in a study to validate a seven day recall questionnaire of physical activity and included 11 overweight boys at a summer camp (Wallace, 1985). Their physical activity was observed by camp counsellors every 15 minutes during the day. At the end of the week of observation the boys were asked to recall their activity for that week. The accuracy of the boys compared with the counsellors was 46% for the type of activity and 75% for the intensity. The authors of this paper claimed that the seven day recall is applicable to children as a summary of their total energy expenditure. The strength of this study is that it uses observed behaviour as a gold standard which may be the best measure available while the weakness of this measure is that it may not be generalizable to free living persons.

A study by Taylor et al (1984) attempted to validate a seven day recall questionnaire with a seven day diary and a Vitalog motion sensor. This study included 12 men from a YMCA cardiac rehabilitation programme and 12 other men from other YMCA courses. The average age of the group was 52 years and they were all white males. The correlations were higher for the weekend activities (range 0.63 to 0.9) than the weekday activities (range 0.39 to 0.75) but the higher values for the weekend correlations may have been due to the lower number of activities recorded in the weekend. The Vitalog correlated with an r value range of 0.38, minus 0.05, 0.91 for moderate, hard, very hard respectively. The most easily remembered activities were those that required
conditioning followed by leisure, home and occupational accordingly. The design strengths of this study are that it uses two different methods for validating the seven day recall. The authors make the point that the participants were not told that the recall would be done at the end of the seven days of recording and consider this to be a strength of the study. The opposite may be true as the one week of daily recording would be likely to assist in the seven day recall at the end of the week. The participants were also highly motivated and the authors suggest that further validation studies be undertaken on average populations.

**Summary:** Recall questionnaires in general and the seven day questionnaire in particular achieve reasonable correlations with other measures of activity and have the advantage of being cheap and easy to administer. This makes the seven day recall a potentially useful instrument for epidemiological studies although more community based assessment is required.

2.1.5. Studies with fitness as a gold standard

The limitation of using fitness as a gold standard are that fitness has a significant genetic component and maximum oxygen uptake does not accurately estimate energy expenditure for low intensity activities (Bouchard, 1990). Moreover fitness testing is not suitable for detecting the mode of activity (Dishman 1988). There is the possibility of bias with physical fitness as the gold standard as participants who exercise regularly may try harder during their
exercise test thereby falsely raising the correlation coefficients. A review of correlations between levels of total leisure time physical activity and physical fitness (as the gold standard) was reported by Lamb and Brodie (1991) in the introduction to a validation study. This review contained thirteen studies with correlations ranging from 0.04 to 0.66.

In their own study, Lamb and Brodie (1991) studied 77 men and 41 women comparing an interviewer administered physical activity questionnaire with a sub-maximal fitness test. The questionnaire related to the previous two weeks of leisure-time activity if those weeks had been typical of the participants regular activity. The interview took between 10 and 20 minutes to complete and the interviewer probed as to the duration and intensity (i.e. occurrence of sweating) of the activities. The univariate correlations were 0.48 for total activity, 0.38 for hard activity and 0.55 for very hard activity. The multiple regression analysis found that very hard activity had a correlation co-efficient of 0.75 for women and 0.84 for men. There were also significant correlations for men only, for body fat, pulse, hard leisure time physical activity and leg power. If these results are verified it may mean that the assessment of physical activity and its benefits needs to assessed separately for each gender. Some of the difficulties with this study are that: participants were only interviewed if the previous two weeks were typical; one outlier was removed from the results and the questionnaire required the interviewer to probe. This latter point would necessitate training
of the interviewer to probe in a consistent fashion. In spite of these disadvantages this method with its probing interviewer may be what is required to improve the validity of questionnaires. This is consistent with the point made by Smith and Morris (1992) where they stated that "... The crux of it is that it is vigorous activity for the given individual." By this they mean that the intensity of physical activity needs to be individualized as this partly circumvents the issue of the genetic contribution to fitness.

In a study evaluating the concurrent and predictive validity for a self reported current activity questionnaire, 95 military officers and 1769 healthy sedentary males were assessed against the same questionnaire administered by an interviewer; maximal oxygen uptake and death by coronary heart disease after seven years (Mundal, 1987). When compared with the gold standard of an interviewer administered questionnaire, the self administered questionnaire had a sensitivity of 46% and a specificity of 96%. Although a small group, the military officers were significantly more active than the sedentary employees. The participants with the lowest level of habitual activity had a significantly higher number of coronary heart disease deaths than those with the highest level of activity. The authors found that the participants with the highest level of activity tended to under-report their activity while some of the inactive group over-estimated their activity. The strengths of this paper are that it uses coronary heart disease deaths as a measure of predictive validity although the authors suggest caution when using the questionnaire to assess
habitual outcome. The weakness of this study is the assumption that the interviewer administered questionnaire is better than the self administered questionnaire although no evidence is offered for this.

In a study evaluating the concurrent validity for a seven day recall of physical activity in community groups and a smaller training study, the recall questionnaire was assessed against maximal oxygen uptake, body fat measurements, food intake and the ability of global questions on physical activity to differentiate between groups (Blair, 1985). In the training study, a randomised controlled trial of the prescription of physical activity for 75 participants, the correlation with maximal oxygen uptake ranged from 0.17 after three months of physical activity to 0.38 at 6 months while for body fat this was -0.11 to -0.50. In the cross-sectional community study involving 1206 women and 1077 men, the seven day recall was able to significantly discriminate between active and inactive persons, change in exercise over a year, manual and other occupations and persons answering yes or no to a global question "do you get enough exercise." Although the analysis was strongly statistically significant in some cases (p <0.05 to < 0.0001) the lack of published correlations makes these results difficult to relate to other studies. The correlation with the 24 hour dietary recall was 0.16 (p<0.001) for men and 0.09 for women (p<0.05) when assessed in a sub-sample of 495 men and 545 women from the community sample. The strength of this study is that the 7 day recall has been compared with a variety of gold standards and
partly conducted in a community sample. The negative aspects are the limitation of maximal oxygen uptake as a gold standard as mentioned above.

In a study evaluating concurrent validity, a 3 day recall questionnaire requiring entries every 15 minutes was compared with fitness, body fat and skin fold thickness (Bouchard, 1983). The study assessed 150 adults and 150 children from Quebec families of French descent; the adults had a mean age of 42.3 years and the children 14.6 years. The three day diary had correlations of 0.7, 0.4 and 0.3 for the physical working capacity at a pulse rate of 150 beats per minute, skin fold thickness and body fat respectively. A repeatability study of 61 persons after 6 to 10 days had correlations of 0.86 to 0.95. The authors concluded that the 3 day diary was a suitable estimate of energy expenditure in population studies. This overstates the case as the subjects were self-selected, had to recall information every 15 minutes and hence liable to change their behaviour as a result of the intrusion into their lives. Also it is not clear if the participants were blinded to the study aims. The strengths of the study were that it involved children and adults, had a high reliability and was conducted over different seasons.

Dishman et al (1988) evaluated the reliability and concurrent validity for a seven day recall activity in college students by comparing the recall questionnaire with a 7 day diary, supervised observation and maximal oxygen uptake. Psychometric aspects such as interest in physical activity were also evaluated and were found
to have no influence on the results. The correlation with the seven day diary was 0.82 and the agreement between the highest and the lowest tertiles was good. However, the correlations with the observed activity were poor, ranging from 0.01 to 0.25. The correlations when compared with the maximal oxygen uptake were 0.61 for total activity and 0.46 for vigorous activity. The repeatability correlations for a re-test after 5 weeks was 0.72. The strengths of this study are that: the 7 day recall has been compared with a variety of gold standards; the participants were unaware they would be completing a 7 day recall at the end of the 7 day diary and they were free living subjects. One negative point was that the participants were educated students. In the maximal testing section the participants were selected so as to have an extreme range of levels of activity, which may have over-estimated the correlation that would occur in community sample. The use of college students also limits the generalizability of the study.

The Minnesota Leisure Time Activity questionnaire is an interviewer administered recall of physical activity during the previous twelve months. The validation of the Minnesota Leisure Time Activity questionnaire used treadmill testing to self-determined point of exhaustion as the gold standard (Taylor, 1978). This study included 175 men randomly selected from a community sample and the detailed results are presented using the time to achieve a pulse of 150 beats per minute as the gold standard. The multiple correlation coefficient for this workload was 0.44. The linear regression coefficients were statistically significant for
heavy and total activity but not significant for moderate and light activity. The authors of this paper claimed an association but qualified this by saying that this did not constitute validity. Although well designed, it is limited in that it only studied men and used fitness as a gold standard.

**Summary:** The use of fitness testing as a gold standard has been a common tool for validating physical activity questionnaires. While reasonable correlations have been shown there is the possibility of bias as highly active participants may be more motivated to perform at high levels on exercise testing. There is also the issue that fitness has a genetic component and fitness testing is not a good measure of physical activity in those who are not genetically endowed for fitness. Fitness testing is unable to validate the mode of activity and the time spent in activities. Finally the main limitation with fitness as a gold standard is the time and cost required to assess this parameter which makes it less attractive as method for large scale epidemiological studies.

**2.1.6. Studies of motion activity sensors**

Large scale integrated activity monitors are devices that can be attached to the body and record body movements. They consist of a mercury switch which closes with movement, each closure is recorded and the results are displayed on a light-emitting diode. These instruments are appealing as they offer the prospect of an objective measure of physical activity. They potentially circumvent
the problem of participants needing to rely on their own memory which can overestimate some aspects of activity and underestimate others. One validation study of these monitors used the Minnesota Leisure Time Activity questionnaire as the gold standard (LaPorte, 1979). This study included 10 male physical education graduates and 10 male graduates from other disciplines. The correlation coefficients with the monitors placed on the trunk was 0.69 (p < 0.01) and 0.43 (p < 0.07) when placed on the ankle. The authors of this paper claimed that the large scale integrated motor activity monitors give an accurate report of activity and that further research was needed to assess the validity in free living individuals.

In another study, large scale integrated activity motor activity monitors were compared using the Paffenbarger Activity Survey as the gold standard (LaPorte, 1983). This study included 76 women with an average age of 61.1 years. The correlation coefficient for the monitors with the Paffenbarger Activity Survey was 0.23 (p<0.05). The correlations with the sub-groups in the Paffenbarger survey were 0.20 with stairs climbed/day, 0.29 for blocks walked per day and 0.58 for sweat episodes. Only the latter two figures were statistically significant at the 0.05 level. The authors of this paper claimed that the large scale integrated activity monitor was an effective and reliable measure of physical activity. While they appear to offer an objective measure of physical activity motion sensors are insensitive to the intensity of exertion and cannot detect the mode of activity (Dishman, 1988).
They are also expensive, have frequent mechanical failures and can potentially alter the behaviour of study participants.

In a study evaluating a mechanical activity counter and a Caltrac electronic accelerometer 50 adults and 30 children were assessed against a gold standard of one hour of observed activity on a Fargo Activity Time sample survey (Klesges, 1985). According to the authors the Fargo scale is a validated instrument. All the correlations were better in the adult group than in the children's group. Although a wide range of correlations were reported the mean intensity ratings ranged from 0.49 to 0.76. The instruments had a correlation of 0.83 with each other in the adult group and 0.42 in the children. The strengths of this study are that the gold standard was a previously validated instrument and the observers were well trained prior to making their observations. The weaknesses of this study are that the observation period was only one hour in duration and the adult participants were first year psychology students and were asked to be active for an hour while being observed; both these aspect make generalization to the wider community limited.

Summary: Activity monitors avoid subjective bias and achieve reasonable correlations with other measures of physical activity. The expense of these devices, frequent breakdowns and inability to detect mode of activity will limit the use of such instruments in large scale epidemiological studies.
2.1.7. Validation of "simple questions"

In an attempt to establish the utility of a simple "sweat" questionnaire Siconolfi et al (1985) performed a validation study using the Paffenbarger Physical Activity Index and a simple sweat question. The sweat question determined the number of times per week that vigorous activity, sufficient to "work up a sweat" was performed. The gold standard was treadmill testing to self-determined point of exhaustion with the highest maximal oxygen uptake being taken as the maximum value. This study included 36 men and 32 women who were hospital employees. The correlation coefficients for the total group, using maximal oxygen uptake as the gold standard, was 0.29 for the physical activity index and 0.46 for the sweat question. The correlation for the sweat test compared with the physical activity index was 0.57. All the correlation coefficients were statistically significant. The authors of this paper concluded that the information from questionnaires was valid and that fitness can be assessed rapidly and simply for epidemiological studies with a simple "sweat" questionnaire. Others have expressed concern about the claims made for rapid assessment methods such as single questions and suggest that they be treated with scepticism until adequately assessed (Smith, 1992).

In another attempt to establish the utility of a simple "sweat" questionnaire Kohl et al (1988) performed a validation study using a mailed questionnaire. One questionnaire asked about the previous 7 days while the other asked about the previous three
months. Both included a simple sweat question which was the same as that used by Siconolfi et al (1985). They also tested an additional activity index specific to running walking and jogging known as the run-walk-jog index. The gold standard was maximal treadmill testing using a modified Balke protocol. This study included 375 men from the Cooper Clinic, a large preventive medicine clinic in Dallas, Texas. The Pearson correlation coefficients, age-adjusted, were 0.51 (p<0.01) for both the sweat frequency and the run-walk-jog index. The co-efficients for the three-month index and the seven-day index were 0.05 (p=0.38) and 0.14 (p=0.03) respectively. When the ability of these measures to predict physical activity was tested using a multiple linear regression model only the sweat frequency and the run-walk-jog index were statistically significant. The authors of this paper conclude that exercise behaviour can be accurately estimated in large populations by using simple questions in a mail survey. The limitation of this study is that the baseline population had been attenders at the clinic and the 12225 participants represented 77% of the original sample and hence the study group was a highly selected population.

The sweat test has also been validated in a community sample in contrast to the other sweat question validation studies which used non-random samples (Washburn, 1990). Washburn et al (1990) wished to assess validity by using high density lipoprotein cholesterol, body mass index and the Harvard Alumni Physical Activity score as the gold standard. This study included 732 men
and women randomly selected from a community sample by random digit dialling. The Pearson correlation coefficients, age and sex-adjusted for sweat hours per week, were 0.23 (p<0.01), -0.02 (P>0.05) and 0.03 (P>0.05) when compared with the Harvard Alumni Physical Activity Score, body mass index and the high density lipoprotein cholesterol. Log transformations of these coefficients increased their magnitude to 0.38 (p<0.01), -0.09 (P<0.05) and 0.11 (P<0.05). The authors of this paper suggest that the utility of self-reported sweat hours may be limited to distinguishing active from inactive subjects in epidemiologic surveys. They qualified this by saying the sweat question would be of limited value in studies in which the physical activity of elderly subjects was being examined as the frequency of exercise-induced sweating would be low. They also thought the sweat question to be too insensitive to detect physical activity during intervention studies. The low correlations in this sweat-test validation probably gives a more realistic estimate than in the selected population of the Kohl et al (1988) study.

A study comparing a single question with a submaximal exercycle test was undertaken in 304 men and women from a communications company (Schechtman, 1991). The single question was "do you currently participate in any regular activity on your own or in a formal class designed to improve or maintain your physical fitness." The results were interpreted as a statistically significant "age adjusted association" with a p value < 0.001. Although the authors implied that this was a valid test for
separating those involved in regular physical activity from those who are not they qualified this by suggesting that this test may not be adequate for assessing changes in physical activity over time. Although a single question is appealing in the context of epidemiological studies this study is limited in that the participants who had the submaximal test were self-selected and the results were not published in a form that allows for comparison with other studies.

**Summary:** Simple questions such as the sweat question appear to be only of value in separating the very active from the inactive. Elderly persons have shown great benefit from pharmacological treatment of hypertension and are likely to be a focus for future research in the blood pressure lowering effects of exercise (Medical Research Council Working Party, 1992); (SHEP Cooperative Research Group, 1991). Simple sweat questions may be inaccurate in that age group. Also better information on the types and intensity of activity in terms of mortality and morbidity is needed. Simple questions will not facilitate this process.

### 2.1.8. A study of multiple methods of measuring physical activity

A comparison of five methods for measuring physical activity in 255 post-menopausal women was done as part of a randomised trial of walking as an intervention to increase bone density (Cauley, 1987). The five methods were the Paffenbarger survey, modified
itors, Baecke survey and a caloric intake. No single measure was considered as a gold standard and each method was compared with the others. The multiple correlations were low (less than 0.15) except where there was some degree of self correlation. A repeatability study in 14 women over 4 weeks had a correlation of 0.76 for total activity. The authors commented on the low values for the correlations even though the measures were meant to be measuring the same activities. The interpretation for these low values was that the instruments may be measuring different facets of activity. The limitation of this study was that it was performed in narrow age-range volunteers for a randomised trial and that the range of activities may have been too small. The strengths of this study are that five methods are compared with each other and in a population with a wider range of ages, higher correlations may have been obtained. In the repeatability study the participants were not aware that they would be repeating the questionnaire and this attention to design detail is good.

2.1.9. Repeatability

Details of the physical activity questionnaire for the Five City Project have been described above (p32) and only repeatability studies, as opposed to validation against a gold standard, have been assessed for this project. The two week test-retest correlations ranged from 0.74 for sleep to 0.075 for moderate activity. Repeatability scores
are usually higher than those for concurrent validity and hence the value of 0.075 is surprisingly low. The next lowest correlation was 0.39 for hard activity which implies that either the result for moderate activity may have been a chance finding or that moderate activity is not easy to distinguish from other activities when using a questionnaire format.

2.1.10. Conclusions
The lack of an accepted gold standard is a major difficulty in validating physical activity questionnaires. Also the majority of the validation studies have been undertaken with small numbers of participants in laboratory settings making the generalizability to the wider population limited. While there are many instruments available to measure physical activity all have some limitation. For large scale studies, cost and the desire not to interfere with the observed behaviour are of paramount importance. For large scale epidemiological studies the simple "sweat" question has merit in dividing groups into two levels of activity but it is not suitable for the elderly and for intervention trials. One of the purposes of the Auckland Heart Study Validation Project was to validate a questionnaire for the Auckland Blood Pressure Control study which was both an intervention study and one in which the majority of participants were elderly. This required a more detailed enquiry into physical activity and thus a questionnaire type of instrument was necessary.
While reviews of physical activity questionnaires and disease outcome have confirmed predictive validity it has also been necessary to establish which aspects of physical activity confer those health benefits (Powell, 1987). The Harvard Alumni Study Physical Activity Index has become a means of comparison for other questionnaires by virtue of it having predicted disease outcomes (Paffenbarger, 1986). Most of the concurrent validation studies achieve Pearson correlation co-efficients of between 0.3 and 0.6. While these seem high a correlation of 0.6 is only explaining 36% of the variance. All of the concurrent validity of questionnaires include a large amount of random error due to the limitations of memory and the variation in physical activity within and between individuals. In studies using physical fitness as a gold standard there is the additional problem of the genetic factors that also influence fitness. Validation of questionnaires using physical fitness will therefore underestimate the true agreement of measures of physical activity.

Despite the measurement problems even the limited questionnaires for measuring physical activity are still predictive of coronary heart disease outcomes. This provides a firm base for further research to assess the components of physical activity that are beneficial to health. For epidemiological studies the Five City questionnaire appears to be the most promising one for present epidemiological purposes but in future more accurate information will need to be obtained. Such questionnaires will require some probing as opposed to the self administered ones to help give better
estimates of the intensity of activity. Hopefully these will be the tools to enable evaluation of lower intensity activities and the relationship with health and disease.
Table 2.1 continued

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Questionnaire type</th>
<th>Gold Standard</th>
<th>N</th>
<th>Age in years</th>
<th>Male/Female</th>
<th>Measures</th>
<th>Range of Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laporte</td>
<td>1979</td>
<td>LSI monitor</td>
<td>3.5 hrs diary</td>
<td>20</td>
<td>?</td>
<td>20/0</td>
<td>Correlation</td>
<td>0.43-0.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12 month recall</td>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7 day recall+diet</td>
<td>76</td>
<td>61</td>
<td>0/76</td>
<td>Correlation</td>
<td>0-0.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sweat question</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
<td>0.58</td>
</tr>
<tr>
<td>Klesges</td>
<td>1985</td>
<td>2 Activity Counters</td>
<td>Observed activity</td>
<td>50/30</td>
<td>21.5</td>
<td>25/25</td>
<td>Correlation</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(average)</td>
<td></td>
</tr>
<tr>
<td>Siconolfi</td>
<td>1985</td>
<td>PPAI questionnaire</td>
<td>VO2 max</td>
<td>68</td>
<td>41</td>
<td>36/32</td>
<td>Correlation</td>
<td>0.29 to 0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sweat question</td>
<td>VO2 max</td>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
<td>0.6</td>
</tr>
<tr>
<td>Kohl</td>
<td>1988</td>
<td>3 month questionnaire</td>
<td>VO2 max</td>
<td>375</td>
<td>47</td>
<td>375/0</td>
<td>Correlation</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 day questionnaire</td>
<td>VO2 max</td>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sweat question</td>
<td>VO2 max</td>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
<td>0.51</td>
</tr>
<tr>
<td>Washburn</td>
<td>1990</td>
<td>Sweat question</td>
<td>Harvard score</td>
<td>732</td>
<td>25-65</td>
<td>Both</td>
<td>Correlation</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HDL cholesterol</td>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Body mass index</td>
<td></td>
<td></td>
<td></td>
<td>Correlation</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schechtman</td>
<td>1991</td>
<td>Single question</td>
<td>Submaximal test</td>
<td>304</td>
<td>37</td>
<td>68/236</td>
<td>Association</td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td>Cauley</td>
<td>1987</td>
<td>5 instruments</td>
<td>5 instruments</td>
<td>255</td>
<td>57</td>
<td>0/255</td>
<td>Correlation</td>
<td>0.01 to 0.48</td>
</tr>
</tbody>
</table>

Key to abbreviations
Caltrac = Caltrac personal activity monitor; an electronic accelerometer
Vitalog = microprocessor to detect heart rate and mercury switch motion sensor.
Submaximal = submaximal exercise testing.
CHD death = coronary heart disease death.
VO2 max = maximal oxygen uptake.
MLTA = Minnesota Leisure Time Activity.
Workload150 = work acheived at a heart rate of 150 beats/minute.
LSI monitor = Large scale integrated activity monitor; mercury switch motion sensor.
PPAI = Paffenbarger Physical Activity Index.
HDL cholesterol = high density lipoprotein cholesterol.
Male/female: The % symbol refers to the percentage of men in the study.
Range of associations: some of these correlations have negative values and only the absolute magnitude is shown.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Questionnaire type</th>
<th>Gold Standard</th>
<th>N</th>
<th>Age in years</th>
<th>Male/Female</th>
<th>Measures</th>
<th>Range of Associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Williams</td>
<td>1989</td>
<td>7 day recall</td>
<td>7 day diary</td>
<td>45</td>
<td>24</td>
<td>21/24</td>
<td>Correlations</td>
<td>0.68 to 0.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Caltrac</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.11 to 0.37</td>
</tr>
<tr>
<td>Wallace</td>
<td>1985</td>
<td>7 day recall</td>
<td>Observed activity</td>
<td>11</td>
<td>12.5</td>
<td>11/0</td>
<td>% accuracy</td>
<td>46% to 75%</td>
</tr>
<tr>
<td>Taylor</td>
<td>1984</td>
<td>7 day recall</td>
<td>7 day diary</td>
<td>30</td>
<td>52</td>
<td>30/0</td>
<td>correlation</td>
<td>0.42 to 0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Vitalog)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.05 to 0.91</td>
</tr>
<tr>
<td>Lamb</td>
<td>1991</td>
<td>14 day recall</td>
<td>Submaximal</td>
<td>118</td>
<td>37.9</td>
<td>77/41</td>
<td>correlation</td>
<td>11.0 to 0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>regression</td>
<td>0.71 to 0.87</td>
</tr>
<tr>
<td>Mundal</td>
<td>1987</td>
<td>Current activity-</td>
<td>Interviewer rating</td>
<td>1864</td>
<td>40-59</td>
<td>1864</td>
<td>p value</td>
<td>p=0.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>self report</td>
<td>CHD death 7 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blair</td>
<td>1985</td>
<td>7 day recall</td>
<td>Running distance;</td>
<td>2283</td>
<td>16-74</td>
<td>53%</td>
<td>Correlation</td>
<td>0.09 to 0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Body fat; VO2 max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Food intake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bouchard</td>
<td>1983</td>
<td>3 day activity</td>
<td>Work at pulse</td>
<td>300</td>
<td>adult 42.3</td>
<td>300</td>
<td>Correlation</td>
<td>0.3 to 0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>report</td>
<td>150/min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Skin fold, body fat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dishman</td>
<td>1988</td>
<td>7 day recall</td>
<td>7 day diary; VO2 max</td>
<td>243</td>
<td>21</td>
<td>Male 55%</td>
<td>Correlation</td>
<td>0.02 to 0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Observation</td>
<td>163</td>
<td></td>
<td>163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor</td>
<td>1978</td>
<td>12 month recall</td>
<td>Fitness</td>
<td>175</td>
<td>36-59</td>
<td>175/0</td>
<td>Correlation</td>
<td>0.44 to 0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MLTA</td>
<td>Workload 150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamb</td>
<td>1991</td>
<td>2 week recall</td>
<td>Submaximal fitness</td>
<td>118</td>
<td>3.8</td>
<td>77/41</td>
<td>Correlation</td>
<td>0.38 to 0.74</td>
</tr>
</tbody>
</table>
Table 2.2  Gold standards as used in the validation of physical activity measures with the advantages and disadvantages of each method.

<table>
<thead>
<tr>
<th>Gold standard</th>
<th>advantages</th>
<th>disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seven day diary</td>
<td>cheap, not reliant on memory, gives information on levels of intensity</td>
<td>may alter behaviour, requires high level of participant co-operation</td>
</tr>
<tr>
<td>Fitness testing</td>
<td>objective</td>
<td>expensive, exercising participants may try harder (bias), potentially dangerous, does not give information on levels of intensity</td>
</tr>
<tr>
<td>Activity monitor</td>
<td>objective</td>
<td>expensive, mechanical problems, measures both small and large body activities, limited use in children</td>
</tr>
<tr>
<td>Heart rate monitor</td>
<td>objective</td>
<td>expensive, mechanical problems, anxiety also raises heart rate</td>
</tr>
<tr>
<td>Observed behaviour</td>
<td>objective, gives information on level of intensity</td>
<td>may alter behaviour, expensive for prolonged use, training required of observers</td>
</tr>
<tr>
<td>Table 2.2 continued</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Food intake</strong></td>
<td>Useful as additional evidence</td>
<td>only useful for total activity usually underestimates total intake</td>
</tr>
<tr>
<td><strong>Questionnaires</strong></td>
<td>Cheap and simple to administer little time an effort for participants</td>
<td>may ask same questions as study (bias) relies on memory</td>
</tr>
</tbody>
</table>
Chapter 3: Methods: Auckland Heart Study Validation Project

3.1.1 Introduction

3.1.2 The Auckland Heart Study

3.1.3 Auckland Heart Study: Three month physical activity recall

3.1.4 Auckland Heart Study: Three month food frequency recall

3.1.5 Auckland Heart Study Validation Project: Self Report Diary

3.1.6 Methods

3.1.7 Statistical methods

3.1.8 Rationale for choosing a one week diary

3.1.9 Summary
3.1.1. Introduction

This chapter describes in detail the methods of the Auckland Heart Study Validation Project. The gold standard for the validation was a seven day self report food intake and physical activity diary. The validation of the dietary questionnaire is only relevant to this thesis in that the total energy input from the dietary diaries was used as a measure of concurrent validity for the total energy output. The validation of the food frequency questionnaire, which was a major part of the validation study, will not be described.

3.1.2. The Auckland Heart Study

The Auckland Heart Study Validation project was a sub-study of the larger Auckland Heart Study, which was a community based case control study of coronary heart disease involving 928 cases and 1568 controls, and conducted in 1986-88. The cases for the Auckland Heart Study came from a community based register and the controls were randomly chosen from the Auckland general electoral roll. Details of the study base and sampling strategy are contained elsewhere (Jackson, 1989). The validation study used a seven day physical activity and food intake diary as a gold standard. (A copy of the food intake and physical activity diary is in appendix A).
Fig 3.1 Overview of Auckland Heart Study Validation Project

Auckland Heart Study Validation Project
Physical activity Questionnaire

validation using physical activity section of diary

Auckland Heart Study Validation Project
Food frequency Questionnaire

validation of food frequency questionnaire using food intake section of the diary
3.1.3. Auckland Heart Study: Three month physical activity recall

The Auckland Heart Study exercise questionnaire documented the average amount of exercise performed over the previous three months (appendix A). An activity was recorded if the participant performed it at least once a week during the three month period before the interview. The exercise questions were based on the: Minnesota Leisure Time Activity questionnaire (Taylor, 1978); Stanford Five City seven day activity questionnaire (Sallis, 1985); Harvard Alumni Study activity questionnaire (Paffenbarger, 1978) and the Health Insurance Plan Study (Shapiro, 1965). No single, previously published, questionnaire was considered suitable for a case-control study and modification of the above questionnaires was required. Short term recall was not appropriate because cases were often interviewed a month after their acute event (usually a myocardial infarction). As next of kin were used as controls for fatal cases detailed questionnaires of the case's previous physical activity was not appropriate. Physical activity questionnaires are usually designed to obtain detailed information about the physical activity of healthy persons and hence may not be appropriate for diseased persons or for retrospective information. Also the questionnaire had to be simple enough for the proxy respondents to complete.

The questionnaire format was similar to that used in the Stanford Five City Project. Questions were asked about hours per
day spent in sleep or rest, moderate, and vigorous activity; light activity was calculated as the remaining hours in the 24 hour day. This approach is efficient because light activities, which are the most frequently performed, do not need to be recorded. The three types of activity considered were leisure (e.g. sports, walking), home (housework, gardening, home maintenance) and occupational. Leisure activity was separated into leisure and home categories to gain a better understanding of the contribution of housework to overall physical activity, most of which is performed by women. Both moderate and vigorous activity are of interest as they appear to be important in protection against coronary heart disease (Powell, 1987) and, since they contribute only a small proportion of daily activity, are likely to be more easily remembered (Taylor, 1978; Washburn, 1986).

3.1.4. Auckland Heart Study:

Three month food frequency recall

The food intake questionnaire used in the Auckland Heart Study was a three month food frequency recall questionnaire. It was a modified version of the Commonwealth Scientific Industrial Research Organization questionnaire. The Auckland Heart Study Validation Project used the 7 day food diary to validate this questionnaire. The food intake section of the diary was being used as a means of concurrent validation for the three month physical activity questionnaire. This enabled the calories from food intake
to be compared with the calories from energy expenditure. The validation of the food frequency questionnaire is not part of this thesis. The food frequency questionnaire is included in this thesis as appendix B.

3.1.5. Auckland Heart Study Validation Project: Seven day Self Report Diary

(1) The Validation Diary: (appendix A)
The diary contained seven single pages for exercise and seven double sided pages for food intake. Thus each day of the week had a single sided page for exercise and a double sided page for food intake. Each physical activity page specified: hours spent asleep and resting; moderate activities and vigorous activities along with examples for each category. Each food intake page specified: breakfast; morning snack; lunch; afternoon snack; evening meal; evening snack; alcoholic drinks and salt intake. Lists of foods commonly consumed at each meal were included on each page as a memory aid. The diary also included instructions and an example of how to fill in the diary. The participants were encouraged to enter details of food consumed in the diary at the time of eating to avoid information loss due to poor recall. The advantage of a prospective diary is that errors due to recall are minimized.
3.1.6 Methods

The aim of the validation study was to re-administer the exercise questionnaire used in the Auckland Heart Study approximately twelve months previously. This was followed by a seven day self-report diary of physical activity and food intake to enable comparisons to be made with the Auckland Heart Study questionnaire and the seven day diary. A random sample of controls from the Auckland Heart Study were selected for the validation study.

After an initial letter of explanation to the participants, a telephone call was made to obtain consent and re-administer the Auckland Heart Study exercise questionnaire by telephone (fig 3.2)
Fig 3.2  Time sequence of Auckland Heart Study Validation Project

- random selection of 1987 AHS controls
- initial letter to participants
- telephone participants; asked for consent; exercise questionnaire administered by phone
- mail self administered food frequency questionnaire (FFQ)
- when FFQ returned diary sent out
- telephone with diary instructions
- when diary returned code food and exercise
- enter physical activity from 3 month recall and diary and food from diary into computer
- perform analysis
In the original study the exercise questionnaire was part of a larger interviewer administered questionnaire. Following the telephone interview for the Auckland Heart Study physical activity questionnaire, the food frequency questionnaire was mailed to the participants. When this was returned, the one week physical activity and food diary was mailed. The participants were asked to start the diary on the following Thursday, if that week was going to be typical of their usual week. The Thursday start was chosen for two reasons; first, it gave the research staff time to contact the participants. This could be quite time consuming as multiple calls were required to make contact, often in the evenings and weekend. Second, there was a concern that some participants would lose interest during the one week recording period and give up filling in the diary. A Thursday start meant that two weekdays were followed by two weekend days which would enable both weekend and weekday information to be obtained in the early part of the diary. In fact, almost all participants who actually started a diary eventually completed it.

A plastic measuring cup and a 5 ml and 15 ml measuring spoon were also mailed to the participants with the diaries. A number of the cups were broken in the mail so participants were asked if they had a cup. Cups were then mailed only to those without them as the majority of participants had measuring cups in their homes. A further telephone call was made to participants a few days after they started the diary. This was discontinued during the study as
most participants had no difficulty with the instructions. The participants were asked to estimate the weight or volume of the food they were consuming. They were not expected to measure everything eaten as this was considered to be too time consuming.

Coding and analysis
The coding of the physical activity questionnaire and the seven day dairy was done by converting the intensity of the physical activity into metabolic equivalents with one metabolic equivalent (MET) being the energy expended by a person while sitting at rest (Taylor, 1978). Moderate activities are defined as activities requiring an energy expenditure of 3-4.9 METs. Vigorous activities are defined as >5 METs. The four categories of activity examined were sleep/rest, light, moderate, and vigorous; the MET values assigned of were 1, 1.5, 4 and 6 respectively. In both the recall questionnaire and the diary, sleep and rest in bed were combined as sleep (i.e. equivalent to 1 MET). The product of the MET value and the time in hours spent in that activity gives an energy expenditure value in Kcal/kg (e.g., brisk walking = 4.5 METs and when performed for 2 hours/day gives a value of 9 Kcal/kg). The MET values, for particular activities, used in this study are described by Wilson et al (1986).

The coding of the food intake diary was done using the University of Otago Food Tables (1984). These are a modification of McCance and Widdowson's tables (Paul, 1978) with New Zealand values added where available. A coding sheet was completed for
each diary. This involved: itemizing every food consumed during each of the seven days; allocating a code number from the Otago Food Tables and recording the weight or volume of food consumed. The weight or volume of each food code item was then summed over the seven days to give a total for the seven days. Each food code and its total for the week was then entered into a computer program written by Professor John Birkbeck using the University of Otago Food Tables. This programme converted the food items into food nutrients. The results were analysed and compared with the: food frequency questionnaire; the seven day physical activity diary; and three month recall.

3.1.7 Statistical methods

(i) Sample size
The sample size calculations indicated that 150 participants would be adequate to ensure significant associations for the range of correlations from 0.3 to 0.6 in the validation study (Donner, 1987). Using the standard levels of significance of alpha = 0.01 and beta = 0.2 the detection of correlations of 0.3 would require 130 participants. A significant correlation of 0.6 was estimated to require 34 participants. A correlation of 0.3 was chosen as this was the minimum value that was thought to be achievable and have face validity. The final decision was based on the logistics of interviewing 150 participants for the validation study.
(II) Analysis

Spearman correlation coefficients, Cohen's kappa score and sensitivity and specificity were the main statistical tools used when comparing the data from the Auckland Heart Study 3 month recall questionnaire with the seven day validation diary. Pearson correlation coefficients were also used and in most situations were similar to the Spearman values. Where there were a large number of tied scores the Spearman method gave the more conservative figures and these are used in the results. Descriptive statistics were calculated using the usual univariate methods. Measures of association such as Cohen's kappa score and sensitivity and specificity were also measured in spite of the majority of other studies using correlation co-efficients. This was done because of the concerns of some authors that correlation co-efficients can give erroneous measures of association (Brennan, 1992). An example of this is where one observer routinely reports twice the value of another yielding a correlation coefficient of 1.0 when in fact there was zero agreement. For validity studies sensitivity and specificity are preferred to the kappa score by some authors as the kappa score does not require a gold standard and its value changes with prevalence (Maclure, 1987). Sensitivity and specificity enable the new measure (i.e. the Auckland Heart Study physical activity questionnaire) to be compared with the gold standard and the values for sensitivity and specificity are essentially stable with changes
in prevalence (Sackett, 1985). A comparison of these measures is made in this thesis as all these values are reported.

3.1.8 Rationale for choosing a one week diary

A physical activity instrument should meet four important criteria (LaPorte, 1985). The instrument should be: valid, reliable, practical and non-reactive (i.e. the instrument must not alter the population or the behaviour it seeks to measure). These criteria can also be applied to food intake measures as there is no single technique which fulfils all these criteria for food intake or physical activity. The choice of a diary format for the physical activity validation was in part influenced by the stronger methodological reasons for choosing the diary format for the food intake part of the validation study. The choice of the diary for the food intake validation is consistent with opinions of major reviewers and investigators in the field. The same cannot be said for the physical activity diary as this area is less developed with regard to validation and diaries are just one of many approaches. As no method has emerged as being clearly better or more popular, the diary was chosen for both the physical activity and the dietary validation.

Combining food intake and physical activity into the same diaries produced efficiency gains. This also enabled an additional measure of concurrent validity to be made by comparing the calorie input, from the food intake, with the calorie output from the physical activity questionnaire. The advantage of a diary is that it
is an immediate form of recall and can be considered a prospective instrument in comparison with the Auckland Heart Study which used a three month retrospective recall for both food intake and physical activity. The errors in the Auckland Heart Study recall questions (i.e. memory limitations, limited food lists and portion sizes) should not occur in the diary thereby avoiding any falsely high agreements. The diary format was logistically feasible for a sample size of 150 and although diary monitoring may alter behaviour this would create random error and result in an underestimate, rather than an overestimate of the correlation. The choice of one week is arbitrary but has merits in terms of content and face validity in that it covers both week and week-end days.

The physical activity diary was designed following the Five-City project format. This involves collecting activity data for the categories of sleep, moderate and vigorous activity. This format is straightforward and covers all activities thereby fulfilling the content validity criteria. This is achieved by measuring only time spent in sleep, moderate and vigorous activity and assuming the rest is light activity. In its original format it has been shown to have acceptable repeatability correlations, ranging from 0.74 for moderate activity to 0.83 for vigorous (Sallis, 1985), and has been used by one other investigator in a diary format (Williams, 1989). As noted above many of the key validation studies have used highly motivated subjects and the Auckland Heart Study Validation diary was designed to ensure good compliance in general population participants. Although more research is needed on the Five-City
recall questionnaire, it has been shown to measure reduction in risk factors for coronary heart disease (Sallis, 1985). Predictive validation against coronary heart disease outcomes in the Five Cities Project has not been possible because of insufficient cardiovascular events. Unfortunately there is no equivalent of the Five City format for food intake.

An alternative strategy was considered in which the Auckland Heart Study physical activity questionnaire would be compared with a questionnaire that had been predictive of coronary heart disease outcomes. The difficulty with this approach would be that errors involved on one recall survey could be the same in one using a similar methodology. The Auckland Heart Study questions were taken from several other studies using questions with predictive validity and falsely high figures could have been obtained as a result of being correlated against itself.

In summary, the use of a diary as a gold standard for validation of the physical activity and food intake questionnaire was a compromise which partly fulfils the four criteria outlined by LaPorte (1985). All the potential techniques for validation involve some compromise and hence the final choice was influenced by practicalities and available resources. The choice of a diary was strongly influenced by it being regarded as the method of choice by workers in the food intake field as well as being a feasible method for the Auckland Heart Study resources. For the physical activity questionnaire, there is no obviously superior method of choice for validation and hence the use of a diary was a reasonable choice. It
also provided administrative efficiency by obtaining two items of information from each study subject using the one instrument. The use of a community based group for validation enables comment on validity as well as generalizability. This is a strong feature of the Auckland Heart Study Validation project as validity studies have usually been done in small, highly motivated populations.

3.1.9 Summary
The Auckland Heart Study Validation Project was a sub-study of the larger Auckland Heart Study. The aim was to validate the 3 month physical activity recall and the 3 month food frequency questionnaire. Only the validation of the 3 month physical activity recall is described in this thesis given the focus on physical activity.

The Auckland Heart Study was a community based case-control study of coronary heart disease involving 928 cases and 1568 controls and conducted in 1986-88. The Validation Project used 152 persons who were controls from 1987 representing 82% of those who were approached. As the Auckland Heart Study controls were randomly chosen from the community the validation study results can be generalized to the community.

The validation instrument was a 7 day diary filled in on a daily basis and used as the gold standard for comparison with the 3 month physical activity recall and food frequency questionnaire. Although a number of measures of physical activity exist, there is
no preferred validation instrument. A diary was chosen as it was suitable as a gold standard for both the physical activity and food intake measurement. The diary contained 7 double sided pages to record food consumed and 7 single sided pages for physical activity. Recording of physical activity was facilitated by using the Five City Study format which measures time spent in moderate activity, vigorous activity and sleep assuming the remaining activity is spent in light activity.
Chapter 4. Results of the Auckland Heart Study Validation Project.

4.1.1. Introduction

4.1.2. Results

4.1.3. Hours spent in work, leisure and home activities

4.1.4. Measures of correlation and agreement
4.1.1. Introduction

This chapter reports the results of the Auckland Heart Study Validation Project for the three month physical activity recall questionnaire. As described in the methods section (chapter 3) the validation instrument was a seven day diary of physical activity and food intake. The results from the validation of the food frequency questionnaire are relevant because the total energy input from the seven day dietary diaries has been used as a measure of concurrent validity for the total energy output from the three month physical activity recall questionnaire. The results from the seven day food diary are reported elsewhere (Arroll, 1991b). The validation of the food frequency questionnaire, which was a major part of the validation study, will not be described in this thesis. Tables 4.1 to 4.6 are included at the end of this chapter from pages 81 to 86.

4.1.2. Results

The Auckland Heart Study Validation project was a sub-study of the larger Auckland Heart Study and details of the latter are described above (3.1.2). For the validation study 208 participants were randomly selected from the Auckland Heart Study controls interviewed in 1987 approximately one year after their initial interview. Of these 22 could not be contacted due to change of address. The remaining 186 people were invited to participate in
the validation study. The seven day diary was completed by 152 people (82% of the contacted sample); 8 people declined to participate and a further 26 did not return the completed diary. Of the 152 who completed the study 75 were women and 77 were men; the average age for the women was 55.7 (SD=7.9) years and for the men 53.2 (SD=9.6) years.

4.1.3. Hours spent in work, leisure and home activities

On average, men spend 7.87, 13.65, 2.25 and 0.23 hours per day in sleep, light, moderate and vigorous activities respectively. Women spend 8.16, 14.33, 1.46 and 0.06 hours per day respectively in the same activities. The proportion of men reporting moderate and vigorous leisure activities was 78% and 16% respectively; 35% and 8% for moderate and vigorous work activities and 95% and 23% for moderate and vigorous home activities respectively. The proportion of women reporting leisure activities was 81% and 3% for moderate and vigorous respectively; 18% and 0% for moderate and vigorous work activities and 99% and 17% for moderate and vigorous home activities. In this population, leisure and home activities constitute the majority of time spent, in hours per week, in moderate and vigorous activity (Table 4.1). Of interest men report more time in activities in the home than women. Much of the home activity that men report was home maintenance and gardening while women reported more time in housekeeping activities.
Table 4.2 shows the hours per day spent in the four levels of activity analysed by age and sex. These results show two trends. First, women report less vigorous activity than men. Second, both men and women in the 35-49 age group report less moderate activities than do the older age groups. This may reflect the limited opportunities for physical activity for men and women during the child rearing period. Also, when people retire, more time is available for work around the house, which would often be of a physical nature. The number of participants in the youngest and oldest age group was very small and no firm conclusions can be made from this data. The results in table 4.3 show the energy expenditure per day spent in the total activity analysed by gender. As expected this shows the same trends as in table 4.2 given that total energy expenditure is the sum of energy in the four categories.

The daily means in Kcal/day for moderate and vigorous physical activity from the three month recall and the seven day diary are shown in table 4.4. The 7 day food intake diary only measures total intake and therefore cannot provide dietary calories consumed in moderate and vigorous activity. There are no significant differences (t test of means) between the recall and the diary method for moderate and total physical activities. There was a significant difference at the vigorous level; for example, the mean vigorous activity in Kcal/day from the seven day diary was $125 \pm 265$ and $67 \pm 207$ from the three month recall. The mean values for total Kcal/day show that values from the three month recall and the seven day diary physical activity are similar. Both
studies use the Stanford Five City Project format in which total activity is the sum of sleep, light, moderate and vigorous activity.

The value for the mean total calorie input from the seven day food diary was substantially lower than those obtained from the other two methods. The seven day food diary appears to underestimate the total daily calorie intake which suggests that participants underestimate their daily food intake. There is no mechanism to account for this phenomenon for food intake studies other than direct observation. In contrast, the Stanford method for measuring physical activity accounts for each hour of the day.

4.1.4. Measures of correlation and agreement
(All correlations in this chapter, from the Auckland Heart Study Validation Project, are statistically significant. There is no significance testing for the kappa score nor the sensitivity and specificity scores as this is not the convention).

The sensitivity and specificity for moderate, vigorous and total activity are presented in Table 4.5. The gold standard for estimating the sensitivity and specificity was the seven day diary. The methods for calculation is described elsewhere (Department of Clinical Epidemiology and Biostatistics, McMaster University II, 1981). In this context the use of the terms sensitivity and specificity denote the ability of the measure of activity to truly represent the actual amount of activity. In this study it is the ability of the 3 month recall to divide persons into the high or low levels of physical activity. It was possible to use a dichotomous variable for vigorous activity as only a proportion of participants performed this level of activity. This was not possible for moderate and total activity as all participants performed some moderate
activity. To obtain a dichotomous variable for both total and moderate activity the median value was taken from the diary and three-month recall. This results in the sensitivity and specificity being identical. The two by two tables and the median values used to dichotomise the variables are shown on Table 4.6. The median values in kcal/day were 3571, 2561 and 2125 for moderate, total activity (for 3 month recall versus the 7 day exercise diary in Kcal/day), and total activity (for 3 month recall versus the 7 day food diary in Kcal/day).

Cohen's Kappa score is a measure of agreement beyond chance and is shown on Table 4.5. The method for describing the kappa score is described elsewhere (Department of Clinical Epidemiology and Biostatistics, McMaster University II, 1980). The Cohen's kappa values for moderate, vigorous, total activity (for 3 month recall versus the 7 day exercise diary in Kcal/day), total activity (for 3 month recall versus the 7 day food diary in Kcal/day) were 0.36, 0.23, 0.62 and 0.22. As with the sensitivity and specificity it was necessary to use a dichotomous variable for moderate and vigorous as all participants performed some moderate and total activity. The two by two tables and median values are shown on Table 4.6.

The Pearson and Spearman correlation coefficients for the two estimates of moderate, vigorous and total daily activity as measured by Kcal/day are shown in Table 4.5. The values for the Spearman correlation coefficients were 0.61, 0.49, 0.86 and 0.41 for moderate, vigorous, total activity (for 3 month recall versus the 7 day exercise diary in Kcal/day), total activity (for 3 month recall
versus the 7 day food diary in Kcal/day). The corresponding values for the Pearson correlation coefficients were 0.60, 0.48, 0.91 respectively. All correlation coefficients are significant at $p < 0.001$. An additional analysis by age and sex was undertaken and the results differed little from the whole sample correlations on Table 4.5. Although all the measures on tables 4.5 and 4.6 have different magnitudes they all have similar gradients of change. In this example there does not seem one statistical measure of agreement that is preferable to the others.
Table 4.1. Mean Time spent in various activities (hours/week ± SD) from 3 month recall questionnaire, by gender.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Level</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leisure</td>
<td>Moderate</td>
<td>2.47 ± 2.03</td>
<td>3.04 ± 3.86</td>
</tr>
<tr>
<td>Leisure</td>
<td>Vigorous</td>
<td>0.19 ± 1.01</td>
<td>0.37 ± 0.92</td>
</tr>
<tr>
<td>Work</td>
<td>Moderate</td>
<td>2.21 ± 6.05</td>
<td>5.76 ± 9.80</td>
</tr>
<tr>
<td>Work</td>
<td>Vigorous</td>
<td>0 ± 0</td>
<td>0.98 ± 3.41</td>
</tr>
<tr>
<td>Home</td>
<td>Moderate</td>
<td>5.52 ± 5.01</td>
<td>6.94 ± 7.40</td>
</tr>
<tr>
<td>Home</td>
<td>Vigorous</td>
<td>0.19 ± 0.47</td>
<td>0.27 ± 0.59</td>
</tr>
</tbody>
</table>
Table 4.2. Hours per day spent in levels of activity ± SD; by age and gender.

<table>
<thead>
<tr>
<th>age/sex(N)</th>
<th>Sleep/rest</th>
<th>light</th>
<th>moderate</th>
<th>vigorous</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-34 men (2)</td>
<td>8.50 ± 0.71</td>
<td>13.62 ± 3.03</td>
<td>1.81 ± 2.37</td>
<td>0.07 ± 0.05</td>
</tr>
<tr>
<td>20-34 women (2)</td>
<td>8.50 ± 0.71</td>
<td>14.68 ± 0.51</td>
<td>0.82 ± 0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>35-49 men (22)</td>
<td>7.79 ± 1.18</td>
<td>13.97 ± 2.42</td>
<td>1.96 ± 1.58</td>
<td>0.28 ± 0.48</td>
</tr>
<tr>
<td>35-49 women (11)</td>
<td>7.83 ± 0.72</td>
<td>14.90 ± 1.1</td>
<td>1.18 ± 0.57</td>
<td>0.08 ± 0.29</td>
</tr>
<tr>
<td>50-64 men (51)</td>
<td>7.83 ± 0.91</td>
<td>13.60 ± 2.18</td>
<td>2.30 ± 1.9</td>
<td>0.16 ± 0.37</td>
</tr>
<tr>
<td>50-64 women (55)</td>
<td>8.32 ± 0.92</td>
<td>14.04 ± 1.53</td>
<td>1.59 ± 1.21</td>
<td>0.05 ± 0.14</td>
</tr>
<tr>
<td>65- men (2)</td>
<td>8.20 ± 0.45</td>
<td>13.43 ± 1.74</td>
<td>2.29 ± 1.43</td>
<td>0.08 ± 0.12</td>
</tr>
<tr>
<td>65- women (5)</td>
<td>7.80 ± 0.84</td>
<td>15.15 ± 1.17</td>
<td>1.02 ± 0.48</td>
<td>0.03 ± 0.06</td>
</tr>
</tbody>
</table>
Table 4.3. Total energy expenditure by age and gender in Kcal/kg/day ± SD

<table>
<thead>
<tr>
<th>Age in years</th>
<th>Men (N)</th>
<th>Women (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-34</td>
<td>36.60 ± 6.03 (2)</td>
<td>33.80 ± 0.86 (2)</td>
</tr>
<tr>
<td>35-49</td>
<td>38.09 ± 4.73 (22)</td>
<td>35.42 ± 2.45 (11)</td>
</tr>
<tr>
<td>50-64</td>
<td>38.77 ± 5.40 (51)</td>
<td>36.04 ± 3.11 (55)</td>
</tr>
<tr>
<td>65-</td>
<td>37.98 ± 3.83 (2)</td>
<td>34.77 ± 1.30 (5)</td>
</tr>
</tbody>
</table>
Table 4.4. Mean values of moderate, vigorous and total physical activity assessed by 3 month recall and 7 day diary in Kcal/day ± SD.

<table>
<thead>
<tr>
<th></th>
<th>3 month physical activity</th>
<th>7 day diary physical activity</th>
<th>7 day diary food intake</th>
</tr>
</thead>
<tbody>
<tr>
<td>moderate</td>
<td>527 ± 448</td>
<td>543 ± 557</td>
<td></td>
</tr>
<tr>
<td>vigorous</td>
<td>67 ± 207</td>
<td>125 ± 268</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>2653 ± 615</td>
<td>2696 ± 712</td>
<td>2249 ± 710</td>
</tr>
</tbody>
</table>
Table 4.5: Comparison between 3 month recall and seven-day diary

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Cohen's kappa</th>
<th>Spearman correlation</th>
<th>Pearson correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate</td>
<td>0.68</td>
<td>0.68</td>
<td>0.36</td>
<td>0.61</td>
<td>0.60</td>
</tr>
<tr>
<td>Vigorous</td>
<td>0.41</td>
<td>0.84</td>
<td>0.23</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td>Total (recall Vs</td>
<td>0.93</td>
<td>0.93</td>
<td>0.62</td>
<td>0.86</td>
<td>0.91</td>
</tr>
<tr>
<td>exercise diary) kcal/day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total (recall Vs</td>
<td>0.61</td>
<td>0.61</td>
<td>0.22</td>
<td>0.41</td>
<td>0.39</td>
</tr>
<tr>
<td>food diary)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

p ≤ 0.01 for all Spearman and Pearson correlations; sensitivity, specificity and kappa scores are not usually expressed with measures of precision.
Table 4.6  Showing 2x2 table cells for the calculation of the sensitivity, specificity and Cohen's kappa score

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Median point diary (gold standard) kcal/day</th>
<th>Median point recall kcal/day</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderate:physical activity</td>
<td>3571</td>
<td>3791</td>
<td>51</td>
<td>24</td>
<td>51</td>
<td>24</td>
<td>0.68</td>
<td>0.68</td>
<td>0.36</td>
</tr>
<tr>
<td>Diary Vs recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vigorous:physical activity</td>
<td>yes/no</td>
<td>yes/no</td>
<td>34</td>
<td>11</td>
<td>49</td>
<td>56</td>
<td>0.41</td>
<td>0.84</td>
<td>0.23</td>
</tr>
<tr>
<td>Diary Vs recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total:physical activity</td>
<td>2561</td>
<td>2558</td>
<td>70</td>
<td>5</td>
<td>5</td>
<td>70</td>
<td>0.93</td>
<td>0.93</td>
<td>0.62</td>
</tr>
<tr>
<td>Diary Vs Recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key:  
a = true positive,  
b = false positive,  
c = false negative,  
d = true negative.
Chapter 5: Discussion

5.1.1. Introduction
5.1.2. Discussion
5.1.3. Comparison with other studies
5.1.1. Introduction
This chapter discusses the results of the Auckland Heart Study Validation project for the three month physical activity recall questionnaire. As described in the methods section (chapter 3) the gold standard for validation was a seven day diary of physical activity and food intake. The results from the validation of the food frequency questionnaire are only relevant to this thesis in that the total energy input from the seven day dietary diaries has been used as a measure of concurrent validity for the total energy output from the three month physical activity recall questionnaire. The discussion and results from the seven day food diary are reported elsewhere (Arroll, 1991b). The validation of the food frequency questionnaire, which was a major part of the validation study, will not be described in this thesis.

5.1.2. Discussion
Reasonable agreement was found between the three month recall questionnaire and the seven day self report diary for moderate and total energy expenditure. In contrast there was a marked underestimation for vigorous activity in the three month recall compared with the diary. The level of agreement for moderate and total activity is encouraging as the validation was performed on a community based population using a prospective validation instrument. The completion rate of 82% is sufficient to enable generalisation to the original Auckland Heart Study control group.
Compared with women, men report more hours spent in leisure, work and home activities at the moderate and vigorous level of activity. No women reported any hours in vigorous work activity. Men reported more time in vigorous work activity than in vigorous leisure activity. However, overall both men and women spend most of their moderate and vigorous activity in non-occupational activities. This confirms the impression that occupations in modern society are mainly sedentary.

To assess the contribution of physical activity in the workplace each participant was asked to rate their own moderate and vigorous occupational activity. The information on occupational activity is open to bias as there are no intensity codes for these activities. This was not the case for home and leisure activities as there were codes available for many of these activities (Wilson, 1986). The subjective impression of the interviewers was that participants may have exaggerated the level of intensity of their work activity to demonstrate demanding nature of their jobs. Examples of this occurred where persons who spent much of day on their feet in an office describing this as moderate or even vigorous activity.

There is a consistency in total energy expenditure by age (table 4.3) for the two age groups 35-49 years and 50-64 years. There is a decline in total energy expenditure for the 65 and over group and although this would be expected the numbers in this group are very small. A similar situation applies to the 30-34 year age group who had unexpectedly low total energy expenditures. For one
of the women this inactivity could be explained by her being in the third trimester of pregnancy.

The mean values for moderate and total activity as assessed by the 3 month recall and seven day diary are consistent. The similarity for the two estimates of total activity is due to the majority of the daily energy expenditure being in sleep and light activities. Most interest is therefore focussed on moderate and vigorous activity as it is these that distinguish between individuals. The mean value for the vigorous component was lower at 67 Kcal/day for the 3 month recall than the 125 kcal/day for the seven day diary. In the three month recall questionnaire only activity done regularly each week was documented whereas in the diary all exercise was recorded. There is the possibility that participants may have performed some additional physical activity during the week of filling in the diary as a consequence of being alerted to physical activity issues. This may partly explain the discrepancy between the recall and the diary for vigorous activity.

The estimate of total activity obtained from the 7 day food diary is 15% lower than the estimate obtained from the two physical activity instruments. This may reflect the Five City format methodology which guarantees to account for 24 hours of activity (i.e. total caloric output) while the food diary will tend to under-estimate the total caloric intake (Livingstone, 1990). The under-estimation is due to participants omitting some food items in their self report diary. Using the Five-City Project format for physical activity measurement such differences are inevitable and
a comparison of food intake with physical activity will at best only provide a rank order of individuals by their physical activity rather than absolute measurements.

Although a significant correlation of 0.61 for moderate physical activity (Table 4.5) is high by population research standards, it only accounts for 36% of the variance of physical activity in the recall questionnaire. The correlation for vigorous physical activity, 0.49, accounts for only 24% of the variance. Apart from the expected variation in human activity, a possible explanation for the amount of variation explained was that the three month recall questionnaire only recorded moderate and vigorous activities performed at least once per week. This eliminated such regular activities as mowing the lawn every two weeks which would appear in the diary on average 50% of the time but never in the recall questionnaire. Other sources of variation could result from the use of a consecutive seven day diary as events such as illness and extremes of weather would have an influence.

The three month recall questionnaire has advantages for a large epidemiological study in that it can be completed in a few minutes and the instructions for using it are straightforward. There is a trade off between the accuracy of a diary and the administrative ease of a three month recall. The act of keeping a diary may also alter the normal pattern of activity thereby potentially reducing the correlation when comparing the diary with the three month recall. Given these influences it is most likely that a correlation of 0.57 for moderate activity and 0.35 for vigorous
activity represent the lower limit of the true correlation between these two instruments. The high correlations for total activity comparing the three month recall with the seven day diary may be an artefact due to both methods using the Five-City format. The smaller Spearman and Pearson correlation co-efficients for the three month recall compared with the seven day food diary may in part explain this although neither instrument is precise.

5.1.3. Comparison with other studies

The results are consistent with those from other investigators who found underestimates for the higher levels of activity and overestimates for the lower levels when comparing retrospective recall instruments with self report (Taylor, 1984); (Klesges, 1985). These misclassifications tended to balance out when the total caloric expenditure was calculated as was the case for our results. The opposite trend was found in another study (Klesges, 1985) and for week-end activities (Taylor, 1984). Such variation in trend may be due to the instrument, the population, or both, and highlights the difficulty in measuring these behaviours.

The correlations in this study are consistent with results from studies using a variety of methodologies (Taylor, 1984); (Williams, 1989); (Sopko, 1984); (Dishman, 1988); (Lamb, 1991); (Wallace, 1985). By obtaining similar results, in our community based sample, we have perhaps answered the concern expressed by
Taylor et al (1984) that their results, from a group of YMCA male volunteers, may not be generalizable.

The correlation of both measures of daily physical activity with daily food intake confirms that the three month recall is a valid measure of physical activity. This is consistent with a study (Sopko, 1984) which found a correlation of 0.5 when comparing maximal oxygen uptake with total food intake. The lower correlation values when using Kcal/kg/day was unexpected as it has been claimed that Kcal/kg/day is a better measure of total physical activity as it is independent of body weight (Sopko, 1984); (Willett, 1986); (Jacobs, 1986).

The sensitivity, specificity and Cohen's kappa score were included as they are considered better measures of association than correlations (Brennan, 1992). Most other studies reported their results in terms of correlations and to facilitate comparison they have been included here. The values of the sensitivity, specificity and Cohen's Kappa score did not alter the rank order of values provided by the correlations when comparing the 3 month recall in comparison with the seven day diary. Sensitivity and specificity are regarded as more appropriate measures of validity than the kappa score (Maclure, 1987). This is because the sensitivity and specificity compare the new questionnaire with the gold standard while the Kappa score does not make that assumption. Thus a high kappa score could occur when both questionnaires were both erroneous in absolute terms but similar in relative terms. Also, as noted in section 3.1.7 kappa scores can change when tested in
different settings while sensitivity and specificity are considered as constant properties.

The issue of just what constitutes validation for physical activity questionnaires is not explicitly discussed in the literature. There is no gold standard in terms of: type of instrument; number of participants; acceptable magnitude of correlation coefficients or their statistical significance. Repeatability is a necessary but not sufficient condition for validation. In general, the published results of physical activity questionnaires show high levels of correlation for repeatability (Williams, 1989); (Sallis, 1985). Validation studies, that involve comparing two different instruments, show a wider range of correlations (LaPorte, 1985); (Taylor, 1978); (Williams, 1989); (Cauley, 1987). The use of food intake as a gold standard has merit in terms of concurrent validity as opposed to the usual practice of using another measure of physical activity. Dietary measurement methodology has many of the same problems as measurement of physical activity in that both suffer from high intra-individual variation. Guidelines for achieving validity are needed in physical activity research as there is no accepted gold standard.

The three month recall questionnaire used in the Auckland Heart Study can be considered to have reasonable validity. Sizeable correlations were achieved with both the physical activity and food intake diary. These correlations will reflect the lower limit of the correlation between the diary and recall methods as a result of intra-individual variation and random error. These results are
similar to other studies of physical activity validity and the validation study has been conducted using the community population for which it was intended. The diary was a prospective instrument and hence errors with it will be largely independent of the three month retrospective recall questionnaire and comparison with food intake provides an additional measure of concurrent validity.

In summary, the Auckland Heart Study three month physical activity recall instrument is measuring physical activity both in general and over the three month recall period. Although valid as measure of physical activity for the purposes of a case-control study its utility and validity in a randomised controlled trial may be limited. Further discussion on the benefits and limitations of this instrument in randomised controlled trials is contained in chapter 10 of this thesis.