CHAPTER 4: METHODS

4.1 Introduction

The purpose of this chapter is to provide a detailed discussion of the econometric methods and models used to examine the relationship between medical care utilisation (MCU) and the level of obesity (BMI) for a sample of the Australian population using individual-level data from the 1995 National Health Survey (NHS). In this chapter the following areas are discussed: (i) structure of the model; (ii) measurement and estimation issues; (iii) variable selection; and (iv) model selection.

4.2 Structure of the Model

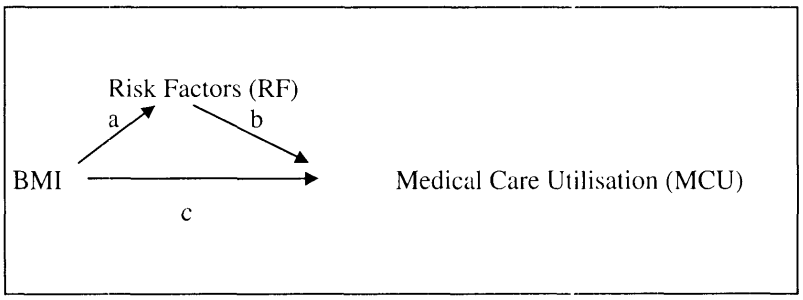
The structure of the econometric models used in this project is based on the supposition that there is a causal relationship between MCU, BMI, and obesity-related risk factors (RF) such as type 2 diabetes mellitus, hypertension, coronary heart disease (CHD), elevated cholesterol levels, depression, and musculoskeletal pain. By definition, causality is implied when "the occurrence of one event is reason enough to expect the production of another" (Heise, 1975, pp. 11-12). The existing research presented in the previous chapters suggests that there *is* a causal relationship between MCU, the level of obesity, and obesity-related risk factors. In particular, the medical literature has demonstrated that obesity leads to a number of serious medical conditions (Pi-Sunyer, 1993; 1996; Sjöström, 1993).

Taking into account this causal relationship, it is expected that individuals with relatively higher levels of BMI will, in turn, have higher MCU. It is postulated that the

relationship runs from BMI to RF to MCU, termed "causal ordering" (Simon, 1987, p. 50). This relationship between MCU, BMI, and RF is the *partial* effect. However, this raises the issue of whether obesity also has an independent impact on MCU.

In this project, it is hypothesised that BMI has zero independent effect. The various relationships are illustrated in Figure 4.1. The partial effect suggested by the medical literature is comprised of *a* and *b*. The total effect is *c*. The working hypotheses are that *a* and *b* (hence a + b) are significant; and that *c* is insignificant.

Figure 4.1: The total and partial effect of BMI on MCU



4.2.1 Recursive Models

In the field of econometrics, it is well known that because of interdependence between the stochastic error term and the explanatory variables, the method of ordinary least squares (OLS) is generally inappropriate for the estimation of a system of simultaneous equations (Green ,1993; Gujarati, 1995; Kennedy, 1992). If OLS is applied in this scenario, the estimators may be biased and inconsistent (Green, 1993; Gujarati, 1995; Kennedy, 1992). Moreover, the bias does not disappear with a relatively large sample size. However, there is a situation in which OLS can be applied even within the context of multiple equation models. This is the case of the recursive model, which is sometimes referred to as the 'triangular' or 'causal' model (Green, 1993; Gujarati, 1995).

To examine this relationship, consider the following three-equation system presented below:

$$Y_{1t} = \beta_{10} + \varphi_{11}X_{1t} + \varphi_{12}X_{2t} + \varepsilon_{1t} \quad (4.1)$$

$$Y_{2t} = \beta_{20} + \beta_{21}Y_{1t} + \varphi_{21}X_{1t} + \varphi_{22}X_{2t} + \varepsilon_{2t} \quad (4.2)$$

$$Y_{3t} = \beta_{30} + \beta_{31}Y_{1t} + \beta_{32}Y_{2t} + \varphi_{31}X_{1t} + \varphi_{32}X_{2t} + \varepsilon_{3t} \quad (4.3)$$

where the Ys are the dependent variables, the Xs are the independent (or explanatory) variables, and the error terms (εs) are not correlated.

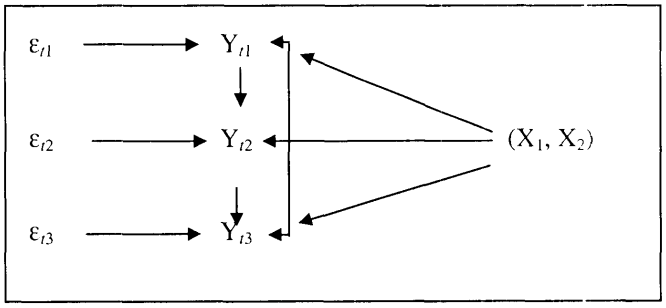
The first equation in this system contains only independent variables on the right-hand side. Since, by assumption, these independent variables are not correlated with the error term (ε_{1t}), this equation satisfies the critical assumption in OLS; that is, there is no correlation between the explanatory variables (X_{1t} and X_{2t}) and the error term (ε_{1t}). Therefore, in this situation, OLS can be applied to the first equation to obtain unbiased estimates.

Consider the second equation, which contains the dependent variable Y_{1t} as an explanatory variable with the X variables. In this instance, OLS can be applied to the second equation provided that Y_{1t} and ε_{2t} are not correlated. This is the case because ε_{1t}, which affects Y_{1t}, is (by assumption) not correlated with ε_{2t}. Therefore, the variable Y_{1t} can be thought of as being 'predetermined' in so far as the variable Y_{2t} is concerned. Therefore, OLS estimation techniques can be applied to the second equation. This

argument can be extended to the third equation because the variables Y_{1t} and Y_{2t} are not correlated with the error term ε_{3t}, and so on.

Therefore, in the recursive system presented above, OLS estimation techniques can be applied to each equation separately. In fact, in this system there is "no simultaneous-equation problem" (Gujarati, 1995, p. 680). Moreover, in this recursive system, Y_{1t} affects Y_{2t}, but Y_{2t} does not affect Y_{1t}. In addition, Y_{1t} and Y_{2t} both affect Y_{3t} without being influenced by Y_{3t}. In more technical terms, each equation exhibits a unilateral causal dependence, hence the name "causal models" (Gujarati, 1995, p. 681). Schematically, the above relationships are presented in the Figure 4.2 below:

Figure 4.2: Representation of a recursive model



Source: Adapted from Gujarati (1995)

4.2.2 The General Model

The relationship between MCU and BMI can be modelled and estimated as a recursive system. First, the recursive system is defined (using a linear specification for convenience at this stage) as follows:

$$RF_i = \alpha_{i0} + \alpha_{i1}CV + \alpha_{i2}BMI + \varepsilon_i \qquad ;i = 1,...,m \qquad (4.4)$$

$$MCU_j = \beta_{j0} + \beta_{j1}CV + \beta_{j2}BMI + \sum_{i=1}^{m} \beta_{j3i}RF_i + \varepsilon_j \qquad ;j = 1,...,n \qquad (4.5)$$

where:

RF_i = obesity-related risk factors;

BMI = body mass index;

CV = a vector of control variables;

MCU_j = medical care utilisation; and

ε_0 and ε_1 are uncorrelated error terms.

In Equations (4.4) and (4.5) there are $i = 1, \ldots, m$ separate risk factors, and $j = 1, \ldots, n$ ways of measuring (indices of) MCU (these issues are considered later). Within this framework it is possible to effectively estimate n models (i.e., one for each measure of MCU), each consisting of m equations (4.4) plus one equation (4.5) (i.e., for one particular j). The working hypotheses that can be tested using (4.4) and (4.5) are presented in Table 4.1 (ignoring subscripts i and j for convenience).

Table 4.1: Principal hypotheses

Hypothesis	*Coefficients Tested*	*Interpretation*
1	$\alpha_2 > 0$	BMI affects RF
2	$\beta_3 > 0$	RF affects MCU
3	$\beta_2 = 0$	BMI has no direct effect on MCU

The first hypothesis states that BMI causes obesity-related risk factors. The second hypothesis states that these risk factors lead to increased MCU (and, hence costs). The

third hypothesis states that BMI has no significant direct (or independent) effect on MCU. Therefore, the recursive system as defined by Equations (4.4) and (4.5) can be estimated using standard OLS techniques. However, as RF is a vector of discrete variables, OLS is inappropriate, and an alternative approach must be employed.

4.2.3 The General Model: Extensions and Further Interpretations

As indicated in the above section, the relationships between BMI, RF, and MCU could, in principle, be modelled as a recursive system. However, there is an alternative approach to estimating this system. To facilitate the following discussion, Equations (4.4) and (4.5) are reproduced below using simplified notation for convenience:

$$RF = \alpha_0 + \alpha_1 CV + \alpha_2 BMI + \varepsilon_0 \tag{4.6}$$

$$MCU = \beta_0 + \beta_1 CV + \beta_2 BMI + \beta_3 RF + \varepsilon_1 \tag{4.7}$$

Since RF in Equation (4.6) is a vector of discrete variables, an alternative approach is to estimate this system as:

$$MCU = A_0 + A_1 CV + A_2 BMI + E_0 \tag{4.8}$$

$$MCU = B_0 + B_1 CV + B_2 BMI + B_3 RF + E_1 \tag{4.9}$$

In this system of Equations, B_2 (4.9) is an estimator for β_2 (4.7), and A_2 (4.8) is an estimator for α_2 (4.6). Therefore, estimating Equation (4.9) tests hypotheses 2 and 3 from Table 4.1. If B_3 (4.9) differs from zero we can then infer that obesity-related risk factors affect (or impact upon) MCU. Furthermore, if B_2 (4.9) is zero or at least close to

zero, this indicates that BMI has little direct effect on MCU. Hypothesis 1 can be tested by estimating Equation (4.8) in addition to (4.9). The estimation of Equation (4.8) can be seen by substituting Equation (4.6) into (4.7) to derive the 'reduced' form:

$$MCU = \beta_0 + \beta_1 CV + \beta_2 BMI + \beta_3(\alpha_0 + \alpha_1 CV + \alpha_2 BMI + \varepsilon_0) + \varepsilon_1 \quad (4.10)$$

which can be rearranged as:

$$MCU = (\alpha_0\beta_3 + \beta_0) + (\alpha_1\beta_3 + \beta_1)CV + (\beta_2 + \beta_3\alpha_2)BMI + (\varepsilon_1 + \beta_3\varepsilon_0) \quad (4.11)$$

The BMI coefficient A_2 (4.8) corresponds to the BMI coefficient in (4.11) – that is, (β_2 + $\beta_3\alpha_2$). If we estimate (4.8), and A_2 differs from B_2 (4.9), then this indicates that BMI and RF are correlated. Taking this argument a step further, it can been seen that if A_2 (4.8) differs from zero, and B_2 is close to zero, then the product of ($\beta_3\alpha_2$) from (4.11) differs from zero. Thus, if B_3 (and therefore β_3) differ from zero, then α_2 cannot equal zero. Hence we can infer that BMI affects RF.

Alternatively, the above relationships can be discussed in terms of path model analysis. For instance, in path model terms, A_2 in (4.8) is being estimated as the measure of the total effect of BMI on MCU, and it is an unbiased measure of the total effect. In addition, B_2 in (4.9) is a measure of the direct effect. Therefore, A_2 less B_2 is referred to as the indirect effect.

Therefore, A_2 (4.8) estimates the coefficient on BMI (4.11) – that is, (β_2 + $\beta_3\alpha_2$). In path model terms, β_2 is the direct effect of BMI on MCU, β_3 is the direct effect of RF on

MCU, and α_2 is the direct effect of BMI on RF. Also, $(\beta_3\alpha_2)$ is referred to as the indirect effect since it is the product of two separate, direct effects. Therefore:

$$A_2 - B_2 = (\beta_2 + \beta_3\alpha_2) - \beta_2 = \beta_3\alpha_2$$

It is hypothesised that β_2 in (4.7) would not be significantly different from zero, and that α_2 in (4.6) would be different from zero. Furthermore, if A_2 in (4.8) is found to be different from zero, it can be inferred that $(\beta_3\alpha_2)$ is different from zero (since B_3 – and therefore β_3 – does not equal zero).

4.2.4 Discussion of the General Model

The previous sections (4.2.2 and 4.2.3) indicate that the relationships between BMI, RF, and MCU can be modelled as a recursive system. However, an alternative approach (which can be derived from the recursive model) may be employed in estimating this system by using Equations (4.8) and (4.9). This approach was discussed in detail in section 4.2.3. Following this discussion, it was also shown how this relationship can be explained in terms of path model analysis.

Consequently, there are several approaches that can be used to interpret and examine the relationships between BMI, RF, and MCU. This, in turn, raises the issue of which approach is preferred. The preferred approach is governed by the objectives of the project, which is to examine not only the relationships between BMI, RF, and MCU but also the policy implications associated with a reduction in BMI.

Taking this into account, the relationships between RF and BMI will be examined in detail according to Equation (4.4). However, the relationship between BMI, RF, and MCU (as represented in Figure 4.1) will be estimated using the approach presented in Equations (4.8) and (4.9). As Equation (4.8) represents the *total* relationship between BMI and MCU it is this particular equation that will form the cornerstone of the subsequent policy analysis.

The simplified models presented here have been used to define the recursive relationships between BMI, RF, and MCU. However, there are many obesity-related risk factors (RF) and a number of ways to define and measure medical care utilisation (MCU). The RF and MCU, and some of the CV variables presented above could be discrete, continuous or in some cases categorical. This is relevant to the particular techniques that can be used to estimate the parameters of the system, which depend crucially on the availability and characteristics of the empirical data. The following sections below discuss measurement issues and selection of variables in the context of the preferred data source, which is the 1995 National Health Survey.

4.3 Measurement and Estimation

Medical research has demonstrated that obesity is a risk factor for a number of serious medical conditions (or diseases). However, whether or not the medical approach to identifying (and measuring) this risk factor and the associated medical conditions are appropriate in the current context is debateable.

This section considers both the conceptual and practical problems associated with the identification and measurement of risk factors for the purpose of this study. It also

considers the measurement of medical care utilisation (MCU), which is to a large extent problematic, given the multi-dimensional nature of medical care.

4.3.1 Measurement of Risk Factors

Definition of Risk and Risk Factors: A Medical Perspective

The term risk, when used in everyday language, denotes that certain activities may be either hazardous or dangerous. For example, to take a risk often means to engage in some potentially dangerous (or risky) activity such as rock climbing or white water rafting. Therefore, to take a risk implies that there is a chance that an unpleasant event (or outcome) may occur.

Within the health science field, the discipline of epidemiology defines the term risk as the *probability that an event will occur* (Unwin et al. 1997). To fully explain the concept of risk as used in the epidemiological and health science literature, it would be beneficial to consider the seminal study by Doll and Hill (1964) who examined the relationship between cigarette smoking and the cause of death in British physicians.

In October 1951, a questionnaire was sent out to 59,600 men and women whose names appeared on the Medical Register in the United Kingdom. This questionnaire was designed to collect information on the smoking habits of these doctors. In total, 40,637 doctors (i.e., a response rate of 68 per cent) returned completed questionnaires. The number of doctors who had died, during the survey period, and their causes of death were recorded (primarily from death certificates). Between November 1, 1955 and October 31, 1961 there were 4,963 recorded deaths. The death rates by smoking status for different causes of death are presented in Table 4.2.

Referring to Table 4.2, the risk of death from all causes for non-smokers was 12.06 per 1,000 persons per year. Individuals who smoked 25 or more cigarettes per day experienced an all cause death rate of 19.67 per 1,000 persons per day. These figures are the *absolute* risk of death among the non-smokers and heavy smokers (i.e., individuals who smoked 25 or more cigarettes per day).

Table 4.2: Smoking status by cause of death (death/1000 persons per year)

Cause of death	*Total population*	*Non-smokers*	*All cigarette smokers*	*Cigarette smokers ≥ 25 a day*
All causes	14.05	12.06	16.32	19.67
Lung cancer	0.65	0.07	1.20	2.23
Coronary heart disease	3.99	3.51	4.57	4.97

Source: Adapted from Doll & Hill (1964)

In epidemiology, the concept of the *relative risk* (RR) is used to estimate the strength of an association between exposure and disease and indicates the likelihood of developing the disease in the exposed group compared to those who are not exposed (Hennekens & Buring, 1987). The relative risk (RR) may be written as follows:

$$RR = \frac{\text{incidence in the group with the exposure}}{\text{incidence in the group without the exposure}}$$

Taking the figures from Table 4.2 for death from lung cancer, the relative risk of all those smoking is:

$$RR = 1.20 / 0.07 = 17.1$$

This means that those smoking were 17 times more likely to die from lung cancer than non-smokers were. In essence, relative risk is a measure of the strength of an

association between an exposure and a disease. Therefore, if the relative risk is equal to 1 the incidence in the two groups is the same. However, if it is greater than 1 then the exposure (i.e., smoking) is associated with an increased incidence in the disease (i.e., lung cancer). An exposure that is positively associated with the occurrence of a disease (such as smoking and lung cancer in this example), is often referred to as a *risk factor for that disease* (Unwin et al. 1997). There is also evidence to suggest that obesity is positively associated with a number of medical conditions. Therefore, in this project, *obesity is a risk factor for a number of medical conditions.*

The following passage in Unwin et al. (1997) provides an historical overview of how the term *risk factor* originated in the medical literature:

> "The idea that different exposures, behaviours and personal attributes influence our risk of developing disease is a very old idea. The concept of 'risk factors', however, comes from modern epidemiology. It has its origins in some of the large prospective studies . . . that were started after the Second World War. The study of the association of smoking behaviour of British doctors with the cause of death is an example of this type of study. Another famous study which helped to establish the concept of 'risk factor' was started in a small town in New England in the United States of America. The town is called Framingham and in the late 1940s male and female residents aged from 30 to 59 years underwent physical examinations, answered questions on personal behaviours, such as smoking, and had blood tests. Over 5000 who were free of coronary heart disease at the time of the examination were re-examined several times over many years to determine who developed coronary heart disease. In this way it was discovered that an increased risk of developing heart disease was associated with smoking, high blood pressure, high serum cholesterol and other factors. These factors were called 'risk factors' for coronary heart disease" (Unwin et al. 1997, pp. 40-41).

From a medical perspective, the idea of identifying *risk factors* for a particular disease is to try to identify factors which may be the causes of the disease and which if modified (or removed) could prevent (or at least minimise) the likelihood of the disease occurring. However, this *association* does not imply *causality* in the medical context, though, further (e.g. scientific) evidence may confirm beliefs about the direction of causality.

The Measurement of Risk Factors: Conceptual and Practical Considerations

There are a number of conceptual and practical issues related to the measurement of risk factors and associated diseases. As stated previously, obesity is a risk factor for a number of medical conditions such as elevated cholesterol levels and heart disease. To examine measurement issues related to risk factors, consider the relationship between cholesterol levels and heart disease will be used as an example.

Consider some medical indicator X, which is continuously measurable, such as cholesterol levels. Cholesterol levels *per se* are not a medical problem (unless levels are significantly elevated) in that it is not associated with pain and suffering or restriction in mobility or function. However, elevated cholesterol levels may lead to (or be a causal factor in the development of) medical condition Y (say, heart disease). Now suppose that Y is equal to either 0 or 1 (that is, you either have heart disease or you do not). Then this relationship can be expressed for the individual as follows:

$$Y = f(X, Z)$$

where Z is a vector of other indicative factors such as lifestyle, age, sex, other medical conditions, and so on.

The concept of a risk factor in medicine, takes the form of specifying some value of X^* of X such that if $X \geq X^*$ then individuals whose cholesterol level is $\geq X^*$ face a significantly greater risk of developing heart disease: cholesterol levels $\geq X^*$ are a risk factor for Y. However, the way in which the level X^* of X is determined by the medical profession is not always clear. For example, in Australia, high cholesterol is defined as

a plasma level 5.5 mmol/L or higher. However, there appears to be no clear answer to the question of how this cholesterol cut off point was defined.

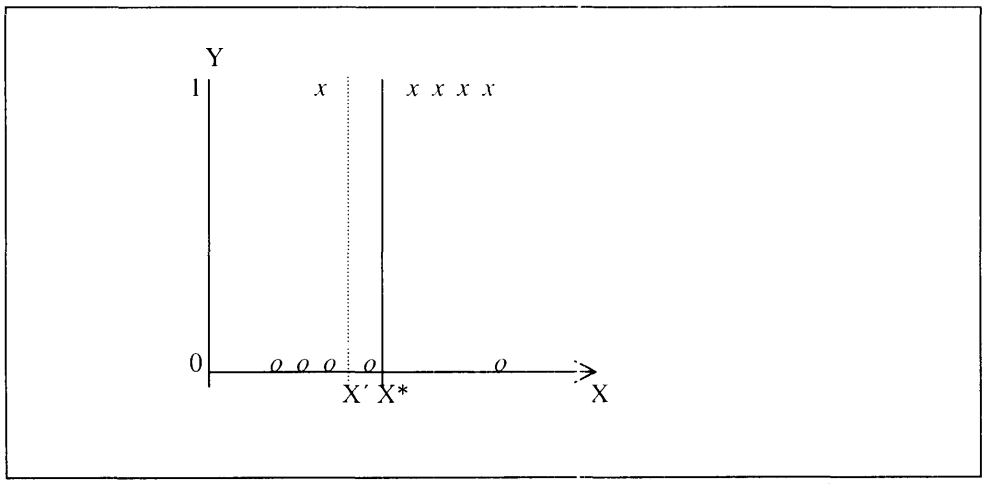
To explore this issue further, consider Figure 4.3, which is a simplified representation of the relationship between cholesterol levels and heart disease for a group of adults aged 20 to 64. The relationship presented in the Figure 4.3 is consistent with the medical evidence that high cholesterol levels are a contributing factor to heart disease.

In Figure 4.3, the Y-axis represents whether or not an individual has heart disease (i.e., 0 = no heart disease; and 1 = heart disease). The X-axis measures cholesterol levels as a continuous variable. High cholesterol levels in the medical sense are represented at X^* (i.e., a cholesterol level $\geq$ 5.5 mmol/L). Plotted in the Figure 4.3 are the cholesterol levels for a group of ten individuals (x = with heart disease; o = without heart disease).

Consider all those individuals who are to the right of cholesterol level X^*. Of these five individuals, four have heart disease and one does not. This indicates that there is an associated probability of 0.8 that individuals with a cholesterol level $\geq X^*$ will have heart disease (other things equal). Now consider that the definition of high cholesterol is revised downwards from X^* to X'. When this is done, all those individuals who are to the right of X' are now considered to have high cholesterol levels. Of these six individuals, four have heart disease and two do not. In this scenario, there is an associated probability of 0.67 that individuals with a cholesterol level $\geq X'$ will have heart disease. This simple example illustrates that the selection of different cholesterol cut off points will result in different probabilities that an individual will have heart disease. In turn, this raises the following issue: why not just measure cholesterol as a

continuous variable? For instance, by measuring cholesterol levels as a continuous variable, probabilities could be estimated for each unit increase using cross sectional data. Using this approach, it would be possible isolate the marginal effect of a one-unit increase in cholesterol levels on the probabi ity that an individual will develop heart disease.

Figure 4.3: Heart disease and cholesterol levels



where: Y = heart disease (yes/no); X = cholesterol levels (continuous); *o* = individuals *without* heart disease; and *x* = individuals *with* heart disease.

From a medical perspective, the association between elevated cholesterol levels and heart disease is supported in published results from long-term prospective clinical trials such as the Lipid Research Clinics Program and the Framingham Heart Study (Kortt & Armstrong, 1998). Large-scale prospective studies, such as Framingham, are specially designed to follow a cohort of individuals over a period of time to see who develops a disease and who does not. Both of these studies provided statistical evidence that elevated cholesterol levels are *associated* with cardiovascular disease. However, for a *causal* relationship to be established further evidence is needed. Initially, other potential explanations have to be excluded such as the possibility of chance, bias, and confounding. Once these possibilities have been ruled out, the criteria for establishing a

causal relationship need to be fulfilled. The criteria for determining a causal relationship are listed in Table 4.3.

Table 4.3: Guidelines for causation

Components	*Interpretation*	*Description*
Temporal relation	Does the cause precede the effect?	This is a crucial relationship. In the current context, elevated cholesterol levels (the cause) must precede heart disease (the effect).
Plausibility	Is the association consistent with other knowledge?	Is the relationship biologically plausible?
Consistency	Have similar results been shown in other studies?	Have several studies produced similar results?
Strength	What is the strength of the association between the cause and effect?	Is there a strong association between the potential cause and the effect as measured by the relative risk?
Dose-response relationship	Is increased exposure to the possible cause associated with the effect?	This relationship occurs when changes in the level of a potential cause are related to changes in the prevalence or incidence of the effect.
Reversibility	Does the removal of a possible cause lead to a reduction of disease risk?	Does the removal of a possible cause result in a reduction of the disease risk.
Study design	Is the evidence based on a strong study design?	Does the evidence come from randomised controlled trials?
Judging the evidence	How many lines of evidence lead to the conclusion?	Causal inference is usually tentative and judgments need to be made on the available evidence.

Source: Adapted from Beaglehole et al. (1993).

These guidelines, presented in Table 4.3, provide a summary and description of the necessary criteria that need to be fulfilled in order to establish causality. For example, this systematic approach has been used by the United States Surgeon General to establish that cigarette smoking caused lung cancer (United States Public Health Service, 1964). If all criteria for a causal relationship are fulfilled and the scientific evidence has been rigorously tested, then in a medical context, a causal relationship is established. However, although the medical profession acknowledges that elevated

cholesterol levels are a risk factor for heart disease, there is debate about the preferred course of treatment for high cholesterol levels.

For example, in the United States, the National Cholesterol Education Program (NCEP) and the American College of Physicians have issued competing guidelines for the treatment of elevated cholesterol levels (Kortt & Armstrong, 1998). These guidelines outline different protocols for the detection, treatment, and management of elevated cholesterol levels. In short, there are differing viewpoints within the medical profession on how to treat this particular risk factor.

Clinical judgment (and experience) may play a crucial role in determining the appropriate course of treatment of risk factors. For example, while there is agreement that elevated cholesterol levels lead to heart disease in adults, this association is controversial in individuals over 65 (Ito, 1996). Ito (1996) recommends that good clinical judgement is required when treating elderly individuals with elevated cholesterol levels. For example, care should be taken when dietary advice is given to elderly individuals, as certain dietary restrictions could potentially contribute to other medical problems. In turn, this raises an important question: "should the focus on cholesterol-lowering therapy interventions be on actually treating elevated cholesterol levels, as opposed to treatment to goal, that is, obtaining a desired cholesterol level according to the NCEP guidelines?" (Kortt & Armstrong, 1998, p. 201). In other words, the crucial point is whether elevated cholesterol levels should be reduced to a specific target (i.e. treatment to goal).

Summarising, from a medical perspective, a set of criteria is used to establish whether a causal link exists between an exposure and disease. However, for particular risk factors such as elevated cholesterol levels, an important issues remains: why (and what is the process behind) selecting cut-off points to classify individuals into different risk groups? There appears to be limited discussion that explicitly addresses this point. In turn, this raises the following issue: to what extent does clinical judgment contribute to actually treating risk factors? While these topics are important to consider, there are other issues in relation to risk factors that should be addressed.

Risk Factors: Further Consideration

The above section addressed the measurement of risk factors (and associated diseases) from both a conceptual and medical perspective. However, a number of other issues should also be considered. In the above example, it is worth noting that elevated cholesterol levels are a modifiable risk factor. Treatment options for elevated cholesterol levels include dietary modification, increased physical activity, or pharmacotherapy. These treatment options can be used to lower cholesterol levels and therefore reduce the risk that an individual will develop heart disease.

Other risk factors may be difficult (or indeed impossible) to modify. For example, being female in and of itself is a risk factor for breast cancer. Also, an individual's environment (or more specifically, geographic location) may also be a risk factor. For instance, living in the state of Queensland may be risk factor for the development of skin cancer. This particular risk factor may be a relatively difficult to modify (especially when one considers the economic constraints that may be associated with

geographic re-location). However, with respect to these risk factors, the main point to stress is that particular risk factors may be difficult (or impossible) to modify.

Measurement of Risk Factors and Diseases: Data Issues

In this project, obesity is defined as a risk factor for the following medical conditions: type 2 diabetes mellitus, hypertension, coronary heart disease (CHD), elevated cholesterol levels, depression, and musculoskeletal pain. To estimate the impact of obesity on these conditions, the best available Australian dataset is the 1995 National Health Survey (NHS). This dataset contains largely self-reported data on medical problems and medical contacts, as well as personal characteristics.

Obesity is measured by the body mass index (BMI). The level of obesity can be measured as a continuous variable using BMI. This differs from the standard medical approach, which is to define obesity in terms of a given BMI value (see section 2.2). BMI (obesity) should be based on measured height and weight. However, the BMI values in the 1995 NHS are calculated on self-reported height and weight. The danger in using self-reported measurements is that there may be a tendency for respondents to overestimate their height and underestimate their body weight. Unfortunately, this type of problem cannot be avoided in self-reported data.

Ideally, the measurement of obesity-associated medical conditions would be identified (using standard diagnostic tests) and measured by trained medical or health care practitioners. This approach would allow for greater accuracy in both the identification and measurement of medical conditions (diseases). It would also assist in reducing the distinct possibility of survey respondents either mis-reporting or under-reporting certain

medical conditions. Moreover, the information collected by medical practitioners could easily be used to measure particular medical conditions (such as type 2 diabetes mellitus, hypertension, and elevated cholesterol levels) as continuous variables. For example, the measurement of blood glucose levels (for type 2 diabetes mellitus), blood pressure (for hypertension), and cholesterol levels would provide extremely useful information. In terms of statistical analysis, the use of continuous variables would form a base from which to conduct detailed statistical analysis.

Unfortunately, the preferred dataset does not provide this level of detail (or accuracy) in relation to the measurement of obesity-associated diseases. As with obesity, these medical conditions are self-reported by survey respondents. In addition, they have been measured (and subsequently recorded) as discrete variables (e.g., the interviewer recorded whether a survey respondent either has or does not have hypertension). Therefore, it is likely that different types of respondents may tend to 'under' and 'over' report certain medical conditions.

In the 1995 NHS, the classification of medical conditions (or diseases) was based on the International Classification of Diseases, 9th Revision (ICD9), but was "modified to suit usage in the NHS" (Australian Bureau of Statistics, 1996, p.124). For example, the classification of coronary heart disease actually consists of several different types of heart disease. This means that particular medical conditions recorded in the NHS may be an aggregation of specific diseases. Although precise classification of medical conditions would be useful, the NHS dataset does provide useful information with which to conduct statistical analysis, even if it is far from ideal.

4.3.2 Measurement of Medical Care Utilisation

The term medical care utilisation (MCU) refers broadly to resource utilisation in the medical sector (e.g., the number of doctor visits or the number of hospital visits). Measurement of MCU poses a number of problems because of its multi-dimensional nature. There is a range of medical care services that can be utilised and each of these services is associated with different levels (or amounts) of medical care usage.

Consider the following characterisation of the issue. There is a range of medical care services, $i = 1, 2, \ldots m$ (e.g., GP consultations, specialist consultations, hospital admissions, and so on) and each type of service is associated with various levels of medical care inputs. Examples include: how much time is spent per doctor visit and how many days are spent in hospital per admission. Hence, each episode of MCU (or services used) i can be represented by a vector of medical inputs: $\{x_{i1}, x_{i2}, x_{i3}, \ldots\}$. In the case of a doctor visit, medical inputs could include: time spent per medical consultation, drugs prescribed, tests administered, and so on. Therefore it is possible in principle to fully describe medical care utilisation, for each individual, as a vector of medical care utilisation (Y) and a matrix of medical inputs (X):

$$\mathbf{X} = \begin{bmatrix} \mathbf{x11} & \mathbf{x12} & \mathbf{x13} & \cdots & \mathbf{x1n} \\ \mathbf{x21} & \mathbf{x22} & \mathbf{x23} & \cdots & \mathbf{x2n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \mathbf{xm1} & \mathbf{xm2} & \mathbf{xm3} & \cdots & \mathbf{xmn} \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} \mathbf{y1} \\ \mathbf{y2} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{ym} \end{bmatrix}$$

Column vector Y represents the different types of medical care services that could be utilised by individuals. For illustrative purposes, consider the variable y_1, which

measures the number of times medical care service type 1 has been used by an individual over a 12-month period (e.g., the number of GP visits). Each utilisation of medical service y_1 is associated with a particular set of inputs, and total inputs used in respect of all services type 1 used is given by: $(x_{11}, x_{12}, x_{13}, \ldots, x_{1n})$. For a single doctor visit, a range of possible medical inputs may include time input by the medical practitioner, time input by nurse (or administrative staff), medical tests administered, quantity of drug *a* prescribed, quantity of drug *b* prescribed, and so on.

In this example, the degree of disaggregation of MCU is, to a large extent, arbitrary. However, the main point to emphasise is that MCU is a multi-dimensional concept. The task here is to represent this comprehensive view of MCU in terms of definitions which are theoretically and empirically manageable. As noted in section 4.2, there are a number of different ways to measure medical care utilisation (MCU). For example, MCU can be aggregated into a single index such as total cost or disaggregated to examine a number of specific types of medical care usage such as 'doctor visits' and 'hospital visits.'

The 1995 NHS contains only limited information on resource utilisation in the medical sector. Specifically, trained interviewers collected information about selected types of health-related actions during the survey period but little about the total medical inputs consumed. Information was collected (and classified) into the following nine categories: (i) hospital episodes; (ii) visits to day clinics; (iii) consultations with doctors; (iv) dental consultations; (v) consultation with other health professionals; (vi) consultation with other persons/organisations; (vii) days absent from work or school;

(viii) other days of reduced activity; and (ix) use of medications. Three of these nine categories are of particular interest in this project (i.e., categories: i, iii, and v,).

The survey items within these three categories can be used as partial indicators of medical care utilisation (MCU). For instance, the 1995 NHS contains a number of survey items for particular types of medical care such as 'doctor visits' (category iii) and 'hospital episodes' (category i). With respect to 'doctor visits', survey respondents were asked: (1) whether consulted a doctor (GP or specialist) in the two weeks prior to the survey; (2) number of doctor consultations in that period; (3) period since last doctor consultation (if not in the last two weeks); and (4) in reference to the most recent doctor visit, the reason(s) for the visit. These data could be used to examine the relationship between MCU (in this case 'doctor visits') and the level of BMI. Furthermore, the data pertaining to 'consultations with other health professionals' and 'hospital episodes' could also be used in a similar fashion. As these components of MCU are likely to be the main cost drivers, it is important to examine their impact separately.

These items on health-related actions are also self-reported. For example, survey respondents were asked whether they had visited a doctor in the two weeks prior to the interview. This is an important point because using two weeks (as opposed to 12 months) as a time frame will reduce the possibility of recall bias.

4.3.3 Estimation Matters

In section 4.2, the relationship between MCU and the level of BMI was described as a recursive system. For convenience, a linear specification was used. However, a linear specification may not be appropriate because of the way in which MCU, BMI, and RF

are measured: the econometric techniques used to estimate these relationships depend on the precise measurement (and definition) of the variables. Moreover, the RF variables (in the 1995 NHS) have been measured as discrete variables. Therefore, for a single attribute, like whether an individual has type 2 diabetes mellitus, the dependent variable RF can only take on two values (0 and 1) defined as:

$RF_i = 1$ if individual i has type 2 diabetes mellitus; and

$RF_i = 0$ otherwise (i.e., individual i does not have type 2 diabetes mellitus).

With respect to Equation (4.4), the dependent RF variables (as recorded and defined in the 1995 NHS) are all discrete. Several statistical techniques may be used to estimate this relationship: linear probability, logistic. or probit models may be used. What cannot be used is OLS.

Equation (4.5) presents the relationship between MCU, the level of obesity (BMI), and risk factors (RF). The MCU variables in the 1995 NHS can be characterised as either discrete or count-based observations. For example, the number of doctor visits in the 1995 NHS are recorded as follows: (i) did you visit a doctor in the last 2 weeks prior to this survey – yes/no?, and (ii) how many times did you visit a doctor in the last 2 weeks prior to this survey – 0, 1, 2, 3 . . . 10 or more times?

Techniques such as logistic regression can be used to estimate the probability of utilising medical care such as doctor visits and are appropriate to use where variables are discrete. On the other hand, the level (or amount) of MCU utilisation can be modelled within a count data framework. Poisson regression (and extensions of this

technique) can be used to model the relationship between the level or the number of times an individual visited a doctor. (This issue is considered further in section 4.4.5 below).

4.3.4 Choosing Control Variables

In this project, the control variables available include age, sex, education, race, family income, employment status, geographic location, and insurance cover. These control variables were included to account for the potential effect that these socio-economic characteristics may have on medical care utilisation. Clinical evidence gives some guidance of the precise selection of control variables to use in estimation, but choice of control variables is ultimately justified by their significance in the estimated equation(s).

Age was selected as a control variable to take into account the likelihood that advancing age will lead to an increase in MCU. Sex was also selected as a control variable to account for the biological differences between men and women which are likely to affect both dependent variables (RF and MCU). These biological differences may potentially have an impact on MCU. The level of education attained by survey respondents was also utilised. It is well known that the levels of education and health status are positively correlated, hence education is likely to have an independent (or explanatory) effect on MCU and (via the lifestyle effects of education) on RF.

An individual's origin was also selected as a control variable to account for the variety of different racial backgrounds in the Australian population. As with sex, differences in racial background (and customs associated with race) may have an impact upon the decision to use medical care and on behaviour affecting the prevalence of risk factors

Sobal and Stunkard (1989) have identified a relationship between socio-economic status (SES) and obesity. This was the main reason why the level of gross personal income was selected as a control variable. It is reasonable to argue that higher levels of income may be associated with greater levels of MCU, other things being equal.

Employment status was also included as a control variable. An individual's employment status may have an independent effect on MCU and could well effect RF.

Geographic location was also selected as a control variable. It is anticipated that those individuals living in metropolitan areas will have greater access to medical services compared to those individuals living in rural areas and hence will have higher levels of MCU.

Health insurance cover was also selected as a control variable to account for differences in the type of health insurance held by the Australian population. Whether an individual has private health insurance may in fact influence MCU. Thus, the type of health insurance cover is likely to have an explanatory impact on MCU (i.e., this variable was included to address the potential problem of moral hazard).

4.4 Variable Selection

The variables, used in the statistical analysis, were selected and extracted from the 1995 National Health Survey (NHS) on CD-ROM. In this project, the study sample consists of 28,376 records (i.e., one record for each individual). In defining the data sample, the following inclusion/exclusion criteria were used: (i) the study sample was limited to individuals between the age of 20 and 64, and (ii) the body mass index (BMI) had to be greater than or equal to 18.5.

Adults under the age of 20 were not included in the analysis because it is reasonable to argue that these individuals could still be developing physically. Also, according to Seidell and Flegal (1997), a healthy (or normal) BMI range is between 18.5 and 24.9. Therefore, people with a BMI value less than 18.5 were excluded from the analysis because the focus of this analysis is on the potential impact of weight loss for high BMI individuals.

The variables selected from the 1995 NHS were classified as either: (i) control variables, (ii) obesity variables, (iii) obesity-related risk factor variables, or (iv) MCU variables. The selection and classification of these variables is based upon the general model presented in section 4.2.2.

4.4.1 CV Measures

Table 4.4 provides details of the control variables used including variable names, codes, frequencies, percentages, and labels. These control variables were selected to account for the potential impact that these characteristics may have on MCU.

Table 4.4: CV measures (n = 28376)

Variable Name	*Code*	*Frequency*	*Per cent*	*Variable Label*
AGE	0	3237	11.4	20-24 years
	1	3535	12.5	25-29 years
	2	3937	13.9	30-34 years
	3	3979	14.0	35-39 years
	4	3733	13.2	40-44 years
	5	3399	12.0	45-49 years
	6	2658	9.4	50-54 years
	7	2131	7.5	55-59 years
	8	1767	6.2	60-64 years
SEX	0	13991	49	Male
	1	14465	51	Female
EDUCATION	0	7143	25.2	No higher qualifications
	1	14040	49.5	NA, inadequately described
	2	277	.0	Higher degree
	3	431	.5	Postgraduate diploma
	4	1504	5.3	Bachelor degree
	5	650	2.3	Undergraduate diploma
	6	786	2.8	Associate diploma
	7	3545	12.5	Skilled/basic vocational
ORIGIN	0	20958	73.9	Australia and New Zealand
	1	2636	9.3	British Isles and Ireland
	2	2374	8.4	Europe
	3	213	.8	Middle East
	4	1397	4.9	Asia
	5	798	2.8	Other
INCOME	0	2183	7.7	$30000-34999
	1	3659	12.9	NA, Don't know / Not Stated
	2	336	1.2	Negative
	3	1852	6.5	$1-4999
	4	3951	13.9	$5000-9999
	5	2311	8.1	$10000-14999
	6	2217	7.8	$15000-19999
	7	3125	11.0	$20000-24999
	8	2843	10.0	$25000-29999
	9	1596	5.6	$35000-39999
	10	1175	4.1	$40000-44999
	11	719	2.5	$45000-49999
	12	679	2.4	$50000-54999
	13	338	1.2	$55000-59999
	14	310	1.1	$60000-64999
	15	176	.6	$65000-69999
	16	174	.6	$70000-74999
	17	732	2.6	$75000 or more
EMPLOYMENT	0	21244	74.9	Wage and salary
	1	1290	4.5	In own business
	2	5842	20.6	Other/NA
GEOGRAPHY	0	16774	59.1	Capital City
	1	3094	10.9	Large/small rural centres
	2	3894	13.7	Other rural area/remote
	3	4614	16.3	ACT/NT
INSURANCE	0	7860	27.1	Does not have private insurance cover
	1	13965	49.2	N/A
	2	6551	23.1	Has private insurance cover

Note: Sample includes individuals aged 20 to 64 with a BMI value ≥ 18.5. In addition, all codes that are equal to zero (0) have been selected as the excluded reference group.

The age variable is classified into nine categories covering the age range 20 to 64. These categories were grouped into five-year intervals ranging from 20-24 (category 1) to 60-64 (category 9). Individuals in the age group 20-24 were selected as the excluded reference group. The sex variable was also selected to distinguish between males and females, with males (category 0) being classified as the excluded reference group. An individual's level of education was classified into eight categories with those individuals who had no higher qualifications selected as the excluded reference group.

An individual's origin was used as a proxy for racial background. Five discrete variables were used to classify an individual's racial background. Individuals who were born in either Australia or New Zealand (category 0) were selected as the excluded reference group.

A control variable for gross personal income was also included. Seventeen discrete variables were used to classify different income ranges. Individuals whose gross personal income was between $30,000-34,000 per annum were selected as the excluded reference group.

Employment status was also selected as a control variable. An individual's employment status was captured by three discrete variables with those individuals who were wage and salary earners selected as reference group.

An individual's geographical location was captured by three discrete variables. Individuals residing in capital cities were chosen as the excluded reference group.

A control variable for private health insurance cover was also selected. An individual's private health insurance cover was captured by two discrete variables. Those individuals who did not have private health insurance were selected as the excluded reference group.

4.4.2 Obesity Measures

The 1995 NHS contains a body mass index variable (BODYMIX) that is calculated on self-reported height and weight. This variable has been classified according to the National Health and Medical Research Council (NHMRC) definition of obesity. Table 4.5 provides details of the body mass index variable contained in the survey.

Table 4.5: Obesity variable (n = 28376)

Variable Name	*NHS BMI Code*	*Frequency*	*Per cent*	*Labels*
• BODYMIX				Body Mass Index
	1	1813	6.4	Less than 20
	2	13347	47.0	20-25
	3	9623	33.9	>25 <30
	4	3593	12.7	≥ 30

Source: Australian Bureau of Statistics (1996).
Note: Sample includes individuals aged 20 to 64 with a BMI value ≥ 18.5.

With respect to Table 4.5, the BODYMIX variable is discrete and four codes were used to group survey respondents into different obesity categories. Using this classification system, 46.6 per cent of individuals (in this sample) were deemed to be either overweight or obese. With this variable, however, it was discovered that there was some mis-classification of BMI values into the above codes. For example, several respondents in the survey were assigned a BMI code of 3 when these individuals should have, in fact, been assigned a BMI code of 2. This example highlights that, in several instances, an incorrect BMI code was assigned to survey respondents.

For this reason, a decision was made to re-calculate BMI using the height and weight variables reported directly in the survey. This re-calculation provides each individual in the sample with a BMI value (as opposed to a code). The reasons for re-calculating body mass index were: (i) the original BMI variable in the survey contained a number of classification errors, (ii) the newly created BMI variable is continuous, and (iii) this new BMI variable can also be re-coded (into discrete variables) according to the WHO classification of obesity. Table 4.6 provides details of the re-calculated BMI variables.

Table 4.6: Re-calculated obesity variables (n = 28376)

Variable Name	*BMI Code*	*Frequency*	*Per cent*	*Labels*
• BMIR				Body Mass Index re-code (discrete)
	1	15295	53.9	≥ 18.5 < 25
	2	9493	33.5	≥ 25 <30
	3	3425	12.1	≥ 30 < 40
	4	163	.6	≥ 40
• BMI	Mean	Min	Max	Body Mass Index (continuous)
	25.30	18.51	52.86	

Source: Australian Bureau of Statistics (1996).
Note: Sample includes individuals aged 20 to 64 with a BMI value ≥ 18.5.

The re-calculated BMI variables summarised in Table 4.6 provide both continuous and discrete measurements of body mass. Both of these variables will be used to examine the association between MCU and BMI for a sample of the Australian population.

4.4.3 RF Measures

The following obesity-related risk factors were selected from the 1995 NHS: type 2 diabetes mellitus, hypertension, coronary heart disease (CHD), elevated cholesterol levels, depression, and musculoskeletal pain. In the dataset, these variables are discrete

(e.g., the interviewer recorded whether a survey respondent either has or does not have type 2 diabetes mellitus). The classification of medical conditions was based on the International Classification of Diseases, 9th Revision, but was "modified to suit usage in the NHS" (Australian Bureau of Statistics, 1996, p.124). Table 4.7 presents the six obesity-related risk factors, the corresponding ICD9 code(s), and NHS re-code.

Table 4.7: Obesity-related risk factors, ICD9 code(s), and NHS re-code(s)

Obesity-Related Risk Factors	*ICD9 code(s)*	*NHS re-code(s)*
Diabetes Mellitus (Type 2)	250.1	78
Hypertension	401-405	72
Coronary Heart Disease (CHD)		
• Heart Disease	391, 393-398, 410-426, 428	82
• Ill-defined signs and symptoms of heart conditions	427, 429	182
Elevated Cholesterol Levels		
• Atherosclerosis	440	15
• High Cholesterol	272.0	108
Depression	300.4, 309, 311	205
Musculoskeletal Pain		
• Rheumatoid arthritis	714	68
• Osteoarthritis	715	69
• Arthritis	711-713, 716	70
• Rheumatism	725,729	89
• Other musculoskeletal disorders	710, 717-721, 723, 726-728, 730-733,739	42

Source: Adapted from Australian Bureau of Statistics (1996).
Note: Selection of obesity-related risk factors was based on a previous study by Kortt (1997).

The selection of the diagnostic codes was based on a previous study by Kortt (1997), in which the selection of the diagnostic codes was reviewed by a team of four clinicians within the Roche Global Pharmacoeconomic Research Department in Palo Alto, California.

In Table 4.7, each medical condition was classified according to an ICD9 code (i.e., the International Classification of Diseases, Edition 9). For example, type 2 diabetes

mellitus has a corresponding ICD9 code (250.1). Subsequently, this particular ICD9 code was re-coded in the 1995 NHS (78). Several medical conditions, such as heart disease (ICD9 codes: 391, 393-398, 410-426, 428), were collapsed into a single NHS code (82).

The 1995 NHS re-codes were used to select those individuals who were classified as having an obesity-related risk factor. Following selection of these risk factors, individuals were coded as either having an obesity-related risk factor or not (i.e., 1 = yes, 0 = no). Table 4.8 provides risk factor variable names, frequencies, mean values, NHS re-codes, and variable labels. It is important to note that the six obesity-related risk factors presented in Table 4.8 are discrete.

Table 4.8: Risk factor variables (n = 28376)

Variable Name	*Frequency*		*Mean*	*NHS Re-code(s)*	*Variable Label*
	RF --Yes	*RF -- No*			
DM_TY2	271	28105	0.010	78	Diabetes Mellitus (type 2)
HYPER	2654	25722	0.094	72	Hypertension
CHOLEST	1774	26602	0.063	15, 108	High Cholesterol
CHD	733	27643	0.026	82, 182	Coronary Heart Disease
DEPRESS	545	27831	0.019	205	Depression
MS_DIS	6595	21781	0.232	68, 69, 70, 89, 42	Musculoskeletal disorders

a Means displayed for discrete variables reflect a proportion.
Note: Sample characteristics include BMI ≥ 18.5 and Aged 20 to 64.

Referring to Table 4.8, type 2 diabetes mellitus, hypertension, high cholesterol levels, and coronary heart disease were selected because there is well-documented evidence that obesity is associated with these diseases (VanItallie, 1985; Garrow, 1991; Colditz, 1992). Depression was also selected as a risk factor. The Swedish Obesity Study (SOS) has indicated that obese adults reported a worse mental health state (Sjöström et al., 1992; Sullivan et al., 1993). Furthermore, obese subjects, and in particular obese women, displayed poorer psychological profiles and mood swings. Based on these

findings, a depression indicator variable was included to examine further the association between obesity and depression. Musculoskeletal disorders were also included because there is evidence linking obesity to a variety of diseases ranging from osteoarthritis to joint pain (Bray, 1985; Colditz, 1992).

However, as noted previously, the obesity-related risk factors in the 1995 NHS are self-reported. The collection of self reported data might lead to certain risk factors being under-reported by survey respondents. Using the study sample, the obesity-related factors were compared with results of other studies to determine if any of these risk factor were under-reported. Table 4.9 presents the number of obesity-related risk factors by BMI for the NHS sample. As shown in Table 4.9, the percentage of individuals who have an obesity-related risk factor increases with the level of BMI. This association is most evident for those individuals who have any one of the following risk factors: hypertension, high cholesterol, or musculoskeletal pain.

Table 4.9: Obesity-related risk factors by BMI (%)

Risk Factors by BMI	*BMI 18.5-24.9*	*BMI 25-29.9*	*BMI 30-39.9*	*BMI ≥40*	*Total*
RISK FACTORS (n)	15295	9493	3425	163	28376
Diabetes Mellitus (Type 2)	51 (0.33)	109 (1.15)	97 (2.83)	14 (8.59)	271 (0.96)
Hypertension	802 (5.24)	1055 (11.11)	742 (21.66)	55 (33.74)	2654 (9.35)
High Cholesterol	613 (4.01)	786 (8.28)	355 (10.36)	20 (12.27)	1774 (6.25)
Heart Disease	305 (1.99)	292 (3.08)	129 (3.77)	7 (4.29)	733 (2.58)
Depression	274 (1.79)	174 (1.83)	90 (2.63)	7 (4.29)	545 (1.92)
Musculoskeletal Pain	2929 (19.15)	2479 (25.11)	1116 (32.58)	71 (43.56)	6595 (23.24)

However, the prevalence of some obesity-related risk factors reported (in the NHS sample) is relatively low compared to other epidemiological estimates. For example, Kluthe & Schubert (1985) report that about 20 per cent of obese individuals also have hypertension. Colditz (1992) also estimates that 26 per cent of obese individuals have

hypertension. In this sample, the comparison rate for hypertension (for those individuals with a BMI ≥ 30) is about 22 per cent, which is consistent with other estimates.

Furthermore, it has been reported that 24-27 per cent of obese individuals also have some type of coronary vascular disease (Colditz, 1992; Wolf & Colditz, 1994; Wolf & Colditz, 1998). In the NHS sample, the comparison rate for CHD is about 3.8 per cent. This discrepancy may, in part, be the result of under-reporting. Therefore, it may be reasonable to argue that the number of risk factors, presented in Table 4.9, is under-reported. This under-reporting of risk factors is a limitation associated with self-reported data.

4.4.5 MCU Measures

The MCU variables are classified as either discrete or count data. Three types of medical care utilisation (MCU) were selected: (i) doctor visits, (ii) other health care professional visits, and (iii) hospital visits. The three discrete variables associated with these types of care are presented in Table 4.10.

In Table 4.10, the ACOL5DOC variable is discrete and is used to indicate whether an individual visited a doctor in the two weeks prior to the interview (1= yes; 0 = no). Included in this definition are consultations by telephone or having a third party (such as a friend or relative) consult a doctor on behalf of the survey respondent. The term doctor includes both general practitioners (GPs) and specialists (such as surgeons, psychiatrists, and so on).

Table 4.10: Discrete MCU variables (n = 28376)

Variable Name	*Mean*	*Min.*	*Max.*	*Variable Label*
Doctor Visits				
• ACOL5DOC	0.22	0	1	Whether consulted any doctors in the 2 weeks prior to interview (yes/no)
Other Health Professionals (OHP)				
• ACOL6OHP	0.108	0	1	Whether visited OHP in the 2 weeks prior to interview (yes/no)
Hospital Visits				
• ACOL1HOSP	0.008	0	1	Whether hospitalised in the 2 weeks prior to interview (yes/no)

Notes: Means displayed for discrete variables reflect a proportion and sample characteristics include BMI ≥ 18.5 and Aged 20 to 64.

The use of other health care professionals (OHP) is also discrete: whether an individual visited an OHP in the two weeks prior to the interview (1 = yes; 0 = no). As defined in the survey, OHP include: chemists, nurses, social workers, psychologists, and so on.

Finally, hospital visits were also selected as a measure of medical care utilisation. The variable ACOL1HOSP is used to identify those individuals who were hospitalised in the two weeks prior to the survey interview (1 = yes; 0 = no).

These discrete MCU variables are also recorded as count data (i.e., data involving counts per time interval) in the 1995 NHS, the number of times an individual utilised medical care was also recorded. The information for these variables are presented in Table 4.11. Included in Table 4.11 is the variable name, the frequency associated with the number of visits, and the variable label.

In Table 4.11, the ANUM5DOC variable contains the number of consultations (with either a GP or specialist) during the survey period. The number of doctor visits recorded in the survey range from no visit (0) to 10 or more visits (10+). In this sample,

6,256 individuals had at least one consultation with a doctor during the survey period. The mean number of doctor visits for this sample is 0.298.

Table 4.11: Count data: MCU variables (n = 28376)

Variable Name	*Frequency*	*Variable Label*
ANUM5DOC		*Number of doctor visits in the last 2 weeks*
0	22120	
1	4755	
2	1098	
3	246	
4	93	
5	37	
6	9	
7	7	
8	0	
9	3	
10+	8	
ANUM6OHP		*Number of OHP visits in last 2 weeks*
0	25298	
1	2241	
2	504	
3	114	
4	117	
5	33	
6	31	
7	8	
8	5	
9	3	
10	10	
11	2	
12	1	
13	1	
14+	8	
ANUM1HOSP		*Number of nights spent in hospital*
0	28162	
1	70	
2	29	
3	26	
4	13	
5	13	
6	13	
7	13	
8	11	
9	6	
10	7	
11	2	
12	1	
13	0	
14	1	
15+	9	

Note: Sample characteristics for all respondents with BMI ≥ 18.5 and Aged 20 to 64.

The ANUM6OHP variable contains the number of consultations with other health professionals during the survey period. The number of OHP visits recorded in the survey range from no visit (0) to 14 or more visits (14+). In this sample, 3,087 people had at least one consultation with another health care professional. The mean number of OHP visits is 0.169.

Finally, the ANUM1HOSP variable lists the total number of nights spent in hospital during the survey period. The total number of nights spent in hospital range from no nights (0) to 15 or more nights (15+). Interestingly, only 214 people were recorded as having spent at least one night in a hospital during the survey period. The mean number of nights spent in hospital for this sample is 0.031.

4.5 Model Selection

This section discusses the selection of the econometric models used to estimate the relationship (or association) between MCU and the level of BMI for a sample of the Australian population. The econometric models, which are presented below, are based on the general model presented and discussed in section 4.2.2.

4.5.1 Modelling the Relationship between RF and BMI

The relationship between MCU and BMI was modelled as a recursive system (refer to Equations 4.4 and 4.5 in section 4.2.2). In this section, the econometric models used to estimate the relationship between RF and BMI is addressed. To facilitate the discussion, a variation of Equation (4.4), is reproduced below. This equation was specified (and subsequently estimated) for each obesity-related risk factor.

$$RF_i = \alpha_{i0} + \sum_{k=1}^{K} \alpha_{i1k} CV_k + \alpha_{i2} BMI + \varepsilon_i \qquad (4.12)$$

where there are $i = 1, \ldots, m$ separate risk factors and K control variables.

As previously noted, obesity is defined as a risk factor for a number of medical conditions. However, some of these medical conditions are, in turn, risk factors for other medical conditions. For example, elevated cholesterol levels are a medical condition and obesity is one of its risk factors, but elevated cholesterol is also a risk factor for heart disease. As a result, the inter-relationship between these risk factors (and obesity) may lead to simultaneity problems in the econometric estimation of Equation (4.12). However, the inter-relationship between these medical conditions can also be modelled as a recursive system (thus eliminating the potential problems associated with simultaneity and identification). These relationships are presented in Table 4.12.

Table 4.12: The inter-relationship between medical conditions

	Possible causes / Associations	→				
Medical Conditions ↓	Diabetes Mellitus (type 2)	CHD	Hypertension	Cholesterol	Depression	Musculoskeletal Pain
Diabetes mellitus (type 2)[1]	✗	✓	✓		✓	
CHD		✗				
Hypertension[2]		✓	✗			
Cholesterol[3]		✓		✗		
Depression					✗	
Musculoskeletal Pain						✗

Notes:
1 – Type 2 diabetes mellitus is a risk factor for CHD (Gerard et al. 1989; Persson, 1995) and hypertension (Gerard et al. 1989; Songer & Zimmet, 1995). Type 2 diabetes mellitus has also been *associated* with anxiety disorders and depression (Hörnquist et al. 1995; Connell et al. 1994; Palinkas et al. 1991).
2 – Hypertension is a risk factor for CHD (Kannel, 1985; Johannesson, 1995; Johannesson et al. 1993).
3 – Elevated cholesterol is a risk factor for CHD (Kannel, 1985; Kortt & Armstrong, 1998).

The six medical conditions listed are also possible risk factors for other medical conditions. For instance, in the elevated cholesterol row, the '✓' indicates that high cholesterol is a causal factor in the development of heart disease. Table 4.12, shows that the inter-relationship between the above risk factors can be modelled as a recursive system. This is based on the following proposition that if there is a '✓' at a_{ij} then a 'blank' at a_{ji} is necessary for there to be no feedback in this system of equations. Since this condition is satisfied in all cases, we can be assured that there is no feedback.

The model selection process was governed by the characteristics of the 1995 NHS dataset. As previously stated, all risk factor variables are discrete (i.e., coded 1 = yes; 0 = no). Statistical techniques may that be used to estimate these relationships include linear probability, logistic, or probit techniques.

Of these three models, the linear probability model has a number of limitations including (i) non-normality of the error term, (ii) heteroscedasticity of the error term, (iii) the distinct possibility that the predicted value of *Y* will lie outside the 0-1 range, and (iv) generally lower R^2 values (e.g., see Gujarati, 1995, p. 552). These limitations mean that the linear probability model will have limited appeal in empirical research. In fact, Griffiths et al. (1993) state that the linear probability model is "generally not recommended for use in practice" (Griffiths et al. 1993, p. 739). As a result, the linear probability model will not be used in this project.

In the current context, either a logistic or probit model could be used to examine the relationship between risk factors and the level of obesity. For illustrative purposes, the logistic model and a variation of this model - the probit model – will be discussed.

The logistic model is defined according to whether the respondent either has a risk factor ($Y = 1$) or does not ($Y = 0$). In this instance, the RF variables consist of six obesity-related risk factors. It is postulated that a set of factors – including age, sex, education, origin, income, employment status, geography, insurance, and BMI – may explain whether or not individuals have a particular risk factor. The relationship between RF and a set of explanatory factors can be expressed as follows:

$$P_i = E(RF_i = 1 \mid CV_k, BMI) = \alpha_{i0} + \sum_{k=1}^{K} \alpha_{i1k} CV_k + \alpha_{i2} BMI \tag{4.13}$$

where the CV_k and BMI are the explanatory variables and $RF_i = 1$ where the individual has risk factor i.

Now consider the following representation of whether or not an individual has a risk factor:

$$P_i = E(Y = 1 \mid CV_k, BMI) = \frac{1}{1 + e^{-(\alpha_{i0} + \sum_{k=1}^{K} \alpha_{i1k} CV_i + \alpha_{i2} BMI)}} \tag{4.14}$$

For ease of exposition, Equation (4.14) can be written as:

$$P_i = \frac{1}{1 + e^{-\theta_i}} \tag{4.15}$$

where:

$$\theta_i = \alpha_{i0} + \sum_{k=1}^{K} \alpha_{i1k} CV_k + \alpha_{i2} BMI \tag{4.16}$$

Equation (4.16) is the cumulative logistic distribution function (Aldrich & Nelson, 1984; Cramer, 1991; Griffiths et al. 1993; Gujarati, 1995). If P_i is the probability of having a risk factor, then (1 - P_i), the probability of not having a risk factor is given by the following expression:

$$1 - P_i = \frac{e^{-\theta_i}}{1 + e^{-\theta_i}} \tag{4.17}$$

Therefore,

$$\frac{P_i}{(1 - P_i)} = \frac{(1/1 + e^{-\theta_i})}{(e^{-\theta_i}/1 + e^{-\theta_i})} = e^{\theta_i} \tag{4.18}$$

Therefore, $P_i / (1 - P_i)$ is the odds ratio in favour of having a risk factor. In other words, $P_i / (1 - P_i)$ is the ratio of the probability that an individual will have a risk factor to the probability that an individual will not have a risk factor. Taking the natural log of Equation (4.18) will result in the following equation:

$$L_i = ln(P_i/1 - P_i) = \alpha_{i0} + \sum_{k=1}^{K} \alpha_{i1k} CV_k + \alpha_{i2} BMI \tag{4.19}$$

where L_i (the log of the odds ratio) is linear in the dependent variables and linear in the parameters (Gujarati, 1995). In the logistic model (L_i) is a linear function of the independent variables, then the probability (P_i) is a non-linear, S-shaped function (Hamilton ,1992). A distinct advantage of the logistic model is that the coefficients can

be interpreted as odds ratios. For discrete variables, the odds ratio equals the anti-logarithm (that is, *e* to the power) of the corresponding coefficient (Cramer, 1991; Hamilton, 1992; Menard, 1995). Odds ratios are used comparatively to examine the "strength of an effect" (Hamilton, 1992, p. 230). However, this is only one approach to interpreting the coefficients in a logistic regression equation. In general, there are three approaches that can be used to interpret logistic coefficients: (i) the log odds (L_i); (ii) ratio of odds ($e^{\theta i}$); or (iii) probabilities [$P_i = 1 / (1 + e^{-\theta i})$].

A variation of the logistic model is the probit (or normit) model. As previously noted, the logistic model uses a cumulative logistic function as shown in Equation (4.14). However, this is not the only cumulative density function (CDF) that can be used (Griffiths et al. 1993). In some applications, the normal CDF has been used (Gujarati, 1995). The model that emerges from the normal CDF is commonly referred to as the probit model. In fact, one could substitute the normal CDF in place of the logistic CDF in Equation (4.14) and could proceed as above. As a result, the formulation of both the logistic and probit models are comparable, with the only real difference being the type of cumulative distribution function used (Griffiths et al. 1993; Gujarati, 1995). In terms of estimation, both models produce similar results. Thus, if both models are theoretically similar and produce similar estimates, which model is preferable in empirical research? Gujarati (1995) provides some insight with respect to this issue:

> ". . . the choice between the two is one of (mathematical) convenience and ready availability of computer programs. On this score, the logit model is generally used in preference to the probit" (Gujarati, 1995, pp. 567-568).

There are also other advantages in using the logistic model over the probit model. First, the logistic coefficients are relatively easier to interpret and can also be expressed as

odds ratios. The logistic model is used here due to the relatively faster computation speed and superior interpretability of the logistic coefficients.

The specification used to estimate the association between type 2 diabetes mellitus (risk factor 1) and the level of BMI for a sample of the Australian population is given by Equation (4.20):[3]

Logistic Model: Probability of type 2 diabetes mellitus (risk factor 1)

$$Li = [ln(P_i/1 - P_i)] = \beta_0 + \beta_1 AGE + \beta_2 SEX + \beta_3 EDUCATION + \beta_4 ORIGIN + \beta_5 INCOME + \beta_6 EMPLOYMENT + \beta_7 GEOGRAPHY + \beta_8 INSURANCE + \beta_9 BMI + \varepsilon_i \quad (4.20)$$

In Equation (4.20), the dependent variable is defined according to whether an individual either has type 2 diabetes mellitus. Thus, Equation (4.20) represents the relationship between the probability that an individual has type 2 diabetes mellitus and a set of explanatory factors. Of particular interest, in Equation (4.20), is the BMI variable and the corresponding logistic coefficient β_9. It is anticipated that the β_9 coefficient will be positive, indicating that there is a statistically significant relationship between BMI and the likelihood that an individual has type 2 diabetes mellitus.

In this equation, the BMI variable can be estimated as either a continuous or discrete variable. If the BMI variable is discrete, then the coefficients in the logistic model can be used to derive an odds ratio which equals the anti-logarithm (that is, e to the power) of the corresponding logistic coefficient.

---

[3] Equation (4.20) is a stylised regression equation in which the β symbols were used instead of α symbols to describe the relationship between RF and BMI to provide a simplified representation of this relationship. The results presented in chapter 5 (and the appendices) use the β symbols for *all* estimated regression coefficients.

In Equation (4.20), the BMI variable is used to classify obesity into four categories. The BMI categories, presented in Table 4.13, are the World Health Organisation (WHO) classification system for obesity. Category (1) is considered to be a 'normal' or 'healthy' BMI range and it is this category that is used as the excluded reference group to compare the different grades of obesity. The classification of the discrete BMI variables is consistent with the following proposition that: "If a qualitative variable has *m* categories, introduce only *m-1* dummy variables" (Gujarati, 1995, p. 504). As there are four obesity categories only three discrete BMI variables are introduced (e.g., BMI1, BMI2, and BMI3).

Table 4.13: Obesity categories and discrete BMI variables

Obesity Categories[a]	*BMI1*	*BMI2*	*BMI3*
(1)BMI 18.5-24.9[b]	0	0	0
(2) BMI 25-29.9	1	0	0
(3) BMI 30-39.9	0	1	0
(4) BMI ≥ 40	0	0	1

a: Categories (2) to (4) are the WHO-endorsed international classification system
b: Reference group

Each discrete BMI variable is being compared to the excluded reference group, namely the 'normal' or 'healthy' BMI range. Therefore, the odds ratio for the three discrete BMI variables may be written as:

$$OR_{\text{BMI25-29.9 versus BMI < 25}} = e^{\beta(\text{on BMI1})}$$

$$OR_{\text{BMI30-39.9 versus BMI < 25}} = e^{\beta(\text{on BMI2})}$$

$$OR_{\text{BMI}\geq 40 \text{ versus BMI < 25}} = e^{\beta(\text{on BMI3})}$$

The above odds ratios are used comparatively to examine the strength of an effect or association (Hamilton, 1992). It is anticipated that the log-odds of type 2 diabetes mellitus will be dependent on the BMI value of an individual. For example, if the odds ratio for BMI3 is equal to 3 then this indicates that individuals in the outermost BMI

category (i.e., a BMI ≥ 40) are 3 times more likely to have type 2 diabetes mellitus compared to individuals within a healthy BMI range. In summary, if the BMI coefficients are statistically significant, then there is evidence of an association between type 2 diabetes mellitus and BMI.

Logistic regression equations for the following obesity-related risk factors – hypertension, CHD, elevated cholesterol levels, depression, and musculoskeletal pain – are summarised in Table 4.14 below. All are the same form as (4.20), with appropriate interpretation.

Table 4.14: Logistic regression equations for risk factors

Equation	*RF (Dependent Variable)*
(4.21)	Hypertension
(4.22)	CHD
(4.23)	Elevated Cholesterol Levels
(4.24)	Depression
(4.25)	Musculoskeletal Pain

Note: The independent (or explanatory) variables are the same as in Equation (4.20).

4.5.2 Modelling the Relationship between MCU, RF, and BMI

The next step was to examine the relationship between MCU, RF, and BMI. Equation (4.26), which is a variation of Equation (4.5), was specified and estimated for each type of medical care utilisation *j*:

$$MCU_j = \beta_{j0} + \sum_{k=1}^{K} \beta_{j1k} CV_k + \beta_{j2} BMI + \sum_{i=1}^{m} \beta_{j3i} RF_i + \varepsilon_j \qquad (4.26)$$

where there are $j = 1, \ldots, n$ ways of measuring MCU, $i = 1, \ldots, m$ separate risk factors and K control variables. As described in section 4.2.3, Equation (4.26) was estimated *with* and *without* the RF variables so in order to examine both the *total* and *partial impact* of BMI on MCU.

The relationship between MCU and BMI was modelled in two stages. The first stage was to estimate the probability (or likelihood) of an individual using medical care. The second stage was to estimate the association between the level (or amount) of MCU and BMI. The components of MCU include doctor visits, visits to other health care professionals, and hospital visits.

In the first stage, a logistic model was defined according to whether a respondent either utilised medical care. The three discrete measures are: (i) whether or not the survey respondent visited a doctor in the two weeks prior to the survey, (ii) whether or not the survey respondent visited another health professional (OHP) in the two weeks prior to the survey, and (iii) whether or not the survey respondent spent a night in hospital in the two weeks prior to the survey. It is postulated that a set of factors – including age, sex, education, race, income level, employment status, insurance, BMI, and RF – may explain the decision of whether or not individuals utilise medical care. For example, the relationship between whether or not the survey respondent visited a doctor and a set of explanatory variables can be expressed in the following stylised regression equation:

Logistic Model: Probability of a doctor visit

$$Li = [\ln(Pi / 1 - Pi)] = \beta_0 + \beta_1 AGE + \beta_2 SEX + \beta_3 EDUCATION + \beta_4 ORIGIN + \beta_5 INCOME + \beta_6 EMPLOYMENT + \beta_7 GEOGRAPHY + \beta_8 INSURANCE + \beta_9 BMI + \beta_{10} RF + \varepsilon_i \quad (4.27)$$

Equation (4.27) highlights the relationship between the probability of a doctor visit and the level of BMI. Initially, Equation (4.27) was estimated *without* the RF variables to assess whether BMI impacts upon MCU (that is, the *total* relationship as discussed in section 4.2.3). Conditional upon the establishment of a relationship between MCU and BMI, it was then examined whether this association is partly operating through the obesity-related risk factors (RF). This is done by estimating Equation (4.27) as presented above. Of particular interest in (4.27) are the BMI coefficients (β_9) and the RF coefficients (β_{10}). If the *introduction* of the risk factor variables (RF) are statistically significant and the BMI variables are statistically insignificant, then this indicates that the impact of BMI is partly exerted through these obesity-related risk factors.

Logistic regression equations for the following types of medical care utilisation (MCU) – other health care professional visits and nights spent in hospital – are summarised in Table 4.15 below. These equations are in the same form as (4.27), with appropriate interpretation.

Table 4.15: Logistic regression equations for MCU

Equation	*MCU (Dependent Variable)*
(4.28)	Visited another health care professional (yes/no)
(4.29)	Spent a night in hospital (yes/no)

Note: The independent (or explanatory) variables are the same as in Equation (4.27).

The second stage was to estimate the association between the *number* of medical services utilised and the levels BMI for survey respondents who: (i) visited a doctor, (ii) visited another health care professional, or (iii) spent a night in hospital. As mentioned previously, the association between the level (or amount) of MCU and BMI can be modelled within a count data framework. Specifically, Poisson regression (and

extensions of this technique) can be employed. However, prior to model specification, it would be helpful to provide an overview of count data models.

As discussed in section 4.4.5 above, the 1995 NHS survey also contains data that involve counts per time interval. For example, as shown in Table 4.11, survey respondents were asked: how many times did you visit a doctor in the two weeks prior to the survey? Responses range from no (0) visits to 10 or more (10+) visits. More generally, the dependent variables presented in Table 4.11 have a number of characteristics that may impact upon the choice of estimation technique. Characteristics of these variables include: (i) a categorical distribution, (ii) an excess number of zeros, and (iii) a relatively long right tail. Several estimation techniques may be suitable for handling data with these particular characteristics (Grootendorst, 1995). Candidate models include the Poisson, negative binomial, and zero inflated count data models. These models are reviewed below.

The Poisson Model

According to Winkelmann (1997), Poisson regression is the benchmark of count data models. In fact, the competing models discussed below are essentially extensions (or variants) of the Poisson regression model. Within the econometrics literature, the Poisson model has been used extensively to model drug utilisation (Grootendorst, 1995), the number of doctor consultations (Cameron & Trivedi, 1985), the discovery of new drugs (Jensen, 1987), and labour mobility (Skrovetz, 1984).

To begin the discussion, consider the probability function of a Poisson-distributed random variable Y,

$$f(y_i) = Pr(Y_i = y_i) = \frac{\lambda_i^{y_i} \, exp(-\lambda_i)}{y_i!}, \; y_i = 0,1,2,... \tag{4.28}$$

and $\lambda_i = exp(x_i\beta)$, where x_i is a vector of independent variables and β is a k x 1 vector of unknown parameters (Caudill & Mixon, 1995; Grootendorst, 1995; Winkelmann, 1997). This particular transformation ensures that the estimated mean of the Poisson model is positive (Grootendorst, 1995). Moreover, this transformation also indicates that the mean and variance of Y_i is equal to λ_i (Caudill & Mixon, 1995; Grootendorst, 1995; Winkelmann, 1997). Cameron and Trivedi (1985) note that the Poisson regression model is based on the following restrictive assumptions. First, it is based on the assumption that "events occur independently over time" (Cameron & Trivedi, 1985, p.31). This is a strong assumption that may not hold for certain types of count data. For instance, the event that individual A saw his/her physician on Friday may not be independent of the event that he/she also saw the physician on Thursday, if both events arise from the occurrence of a single medical condition. Secondly, the assumption that the conditional mean and variance of y_i given x_i are equal may also be a restrictive assumption failing to account for overdispersion – that is, the variance exceeds the mean (Cameron & Trivedi, 1985; Grootendorst, 1995; Winkelmann, 1997).

In addition to the above limitations, there is another important point worth noting. As presented in Table 4.11, the distribution of these variables has been right censored. For example, consider the number of doctor visits recorded during the survey period. Any observation greater than 10 is masked by being labelled a 10 (Green, 1993). Likewise,

the number of other health care professional visits and number of nights spent in hospital are also masked by the labels 14 and 15, respectively. Data with these particular characteristics can be modelled within a Poisson framework (Caudill & Mixon, 1995; Green, 1993). The Poisson model with right censoring is the same as (4.28) above except that for some integer C, all values of y greater than or equal to C are reported as C (Green, 1993).[4] The formulation of this model follows the exposition by Caudill & Mixon (1995) and Green (1993). First, a *latent* variable[5], Y, is defined as an underlying Poisson variable:

$$\text{Prob}(Y^* = y) = \frac{\lambda^y \exp(-\lambda)}{y!} \tag{4.28a}$$

where $\lambda = exp(x\beta)$

The *observed* variable is therefore:

$$y = \text{Min}[y^*, C] \tag{4.29}$$

Thus,

$$\text{Prob}[y = j] = \text{Prob}[y = j] = \text{if } y < C \tag{4.30}$$

and;

---

[4] In essence, this a Tobit model for count data (e.g., see Green, 1993).
[5] In the current context, the latent variable can be thought of as desired MCU.

$$\text{Prob}(y = C) = \text{Prob}(y \geq C)$$
$$= (1 - \text{Prob}[y < C])$$

$$= 1 - \sum_{j=0}^{C-1} \text{Prob}(y^* = j) \tag{4.31}$$

This shows how censored count data can be handled within a Poisson framework. However, owing to the restrictions associated with the Poisson model, other specifications may, in fact, be more appropriate.

The Negative Binomial Model

Certain types of count data are characterised by overdispersion (i.e., the conditional variance exceeds the conditional mean). Overdispersion may arise due to unobservable individual heterogeneity in the dependent variable (Cameron & Trivedi, 1985; Grootendorst, 1995). The negative binomial model arises if this inter-person heterogeneity is modelled using the gamma probability distribution (Cameron & Trivedi, 1985; Grootendorst, 1995; Winklemann, 1997). Specifically, the density of the negative binomial is derived by including an error term to the conditional mean of the Poisson (Grootendorst, 1995):

$$\lambda_i = exp(x_i\beta) + \varepsilon \tag{4.32}$$

where $exp(\varepsilon)$ follows a gamma distribution with mean one and variance α. Moreover, substituting (4.32) into (4.28a) above and integrating ε out of the expression will result in the negative binominal density:

$$f(y_i) = \text{Prob}(Y_i = y_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta) y_i!} u_i \theta (1 - u_i) y_i,$$

$$y_i = 0, 1, 2, \ldots \tag{4.33}$$

where:

$$u_i = \frac{\theta}{\theta + \lambda_i}, \theta = \frac{1}{\alpha}, \Gamma(\cdot) = \text{gamma function}$$

Therefore, the introduction of the parameter α allows the mean to differ from the variance. Ultimately, this extension of the Poisson model can be used to model data characterised by overdispersion. With the availability of specialised econometric programs such as LIMDEP by William Greene, the estimation of the negative binomial model is relatively straightforward. More importantly, testing for overdispersion "conveniently reduces to a *t*-test on the sign ficance of the estimated value of α" (Grootendorst, 1995, p.185). That is:

H_o: $\alpha = 0$ (the Poisson model is favoured)

H_a: $\alpha > 0$ (the negative binomial model is favoured)

It is also important to note that the negative binomial model with censoring can be obtained by altering the functional form of the probability (this is analogous to the censored Poisson model discussed above).

The Zero Inflated Count Data Models

In some cases, the Poisson and negative binomial regression models may not accurately assign the probability to the outcome $Y = 0$ (Green, 1994; Winkelmann, 1997). For example, in a survey of the number of times an individual consulted a doctor, the outcome for $Y = 0$ may arise; at another time, that same individual might choose $Y = j > 0$. This is a subtle point as the response 0 is distinct from the response '0 and none planned.' Consequently, zero inflated Poisson or negative binomial models can be used to address this issue (Green, 1994; Grootendorst, 1995; Winklemann, 1997). The underlying theory associated with these models follows the exposition by Winklemann (1997). To begin with, consider a binary variable c_i that allows for a separate treatment of zeros and positive outcomes. Following on from this let:

$$y_i = \begin{cases} 0 \text{ if } c_i = 0 \\ y_i^* \text{ if } c_i = 1 \end{cases} \tag{4.34}$$

Thus, if the probability of $c_i = 1$ is denoted by p_i, then the associated probability function of y_i is:

$$g(y_i) = (1 - p_i)^{1-e_i} + p_i f(y_i) \tag{4.35}$$

With respect to zero inflated count data models "there are two types of zeros: one type is obtained as $c_i = 0$; the other as $c_i = 1$ and $y_i^* = 0$" (Winkelmann, 1997, p.107). However, the selection of which of these two models is more appropriate will ultimately dependent on the application and characteristics of the data. However, in practice, it is possible to estimate both these models and then examine which is the preferred model.

Selecting the Preferred Count Data Model

As reviewed above, there are a number of candidate models to select from. To facilitate the discussion, these competing models are summarised in 4.16 below. Ultimately, selection of the preferred model is, to a large extent, dependent upon the characteristics of the dataset and the empirical results obtained. In principle, it would be feasible to estimate all the above models.

Table 4.16: Candidate count data models

Models
(1) Poisson & censored Poisson
(2) Negative binomial & censored negative binomial
(3) Zero inflated Poisson (ZIP)
(4) Zero inflated negative binomial (ZIB)

However, from a practical viewpoint it would preferable to estimate a base model from which to compare alternative models. Thus, a decision was made to estimate the Poisson and censored Poisson models first. This decision was based on the following reasons: (i) the Poisson model is the benchmark for count data models (Winkelmann, 1997); and (ii) in the applied econometric literature, the Poisson model has been used as a starting point in analysing count data models. In addition, the other models presented in Table 4.16 are essentially extensions of the Poisson model. It is entirely possible that owing to the characteristics of the dataset, the estimation of alternative models may, in fact, provide little, if any, improvement over the Poisson model. This is, however, an empirical question. Therefore, as a starting point, the Poisson and censored Poisson regression models were specified and estimated for the three different types of medical care utilisation. Poisson and censored Poisson equations for the three types of medical care utilisation are summarised in Table 4.17.

Table 4.17: Poisson and censored Poisson equations for MCU

Equation	*MCU (Dependent Variable)*
Poisson	
(4.36)	Number of doctor visits
(4.37)	Number of visits to other health care professional
(4.38)	Number of nights spent in hospital
Censored Poisson	
(4.36a)	Number of doctor visits
(4.37a)	Number of visits to other health care professional
(4.38a)	Number of nights spent in hospital

Following the estimation of the above equations, the alternative count data models listed in Table 4.16 were also investigated. The purpose of this was to examine whether extension of the Poisson model was warranted.

4.6 Summary

This chapter has provided a discussion of the econometric methods and models used to examine the relationship between MCU, BMI, and RF for a sample of the Australia population. The following chapter presents (and discusses) the results from these econometric models, with particular reference given to the working hypotheses presented in Table 4.1.

CHAPTER 5: RESULTS

5.1 Introduction

This chapter presents the estimates from the econometric models discussed in chapter 4. The econometric analysis of the 1995 NHS dataset was performed using the statistical software packages SPSS (Version 8.0) and LIMDEP (Version 7). SPSS was used to estimate the logistic regression equations while LIMDEP was employed to estimate the count data models. The results reported in this chapter include: (i) descriptive statistics of the sample by the level of BMI; (ii) logistic estimates of the relationship between RF and the level of BMI, (iii) logistic estimates of the relationship between MCU and the level of BMI, and (iv) Poisson estimates for the relationship between the *level* of MCU and BMI. The results are presented and discussed with particular reference to the working hypotheses presented in chapter 4 (refer to Table 4.1). The final section in this chapter provides a summary of the principal findings.

5.2 Characteristics of the Dataset

The data used in this analysis were extracted from the 1995 Australian National Health Survey (NHS) on CD-ROM. This sample includes 28,376 unique records (i.e., one record for each individual). The results presented in this section are for adults (age between 20 and 64 years) with a BMI value greater than or equal to 18.5. Table 5.1 highlights the characteristics of the 28,376 individuals in the sample by BMI (as a percentage of the total). Included in Table 5.1 are the control and obesity-related risk factor variables.

Table 5.1: Characteristics of the 28,376 individuals in the sample by BMI (% of total)

Characteristics	*BMI 18.5-24.9*	*BMI 25-29.9*	*BMI 30-39.9*	*BMI ≥ 40*	*Total*
AGE					
20-24 years	2315 (71.5)	707 (21.8)	210 (6.5)	5 (0.2)	3237
25-29 years	2186 (61.8)	1007 (28.5)	324 (9.2)	18 (0.5)	3535
30-34 years	2250 (57.2)	1222 (31)	442 (11.2)	23 (0.6)	3937
35-39 years	2235 (56.2)	1292 (32.5)	427 (10.7)	25 (0.6)	3979
40-44 years	1953 (52.3)	1309 (35.1)	450 (12.1)	21 (0.6)	3733
45-49 years	1590 (46.8)	1280 (37.7)	504 (14.8)	25 (0.7)	3399
50-54 years	1146 (43.1)	1070 (40.3)	425 (16.0)	17 (0.6)	2658
55-59 years	877 (41.2)	879 (41.2)	354 (16.6)	21 (0.1)	2131
60-64 years	743 (42.0)	727 (41.1)	289 (16.4)	8 (0.5)	1767
SEX					
Female	8671 (62.3)	3432 (24.7)	1677 (12.1)	131 (0.9)	13911
Male	6624 (45.8)	6061 (41.9)	1748 (12.1)	32 (0.2)	14465
HIGHER EDUCATION					
No higher qualifications	3721 (52.1)	2398 (33.6)	966 (13.5)	58 (0.8)	7143
NA, inadequately described	7588 (54.0)	4687 (33.4)	1681 (12.0)	84 (0.6)	14040
Higher degree	156 (56.3)	105 (37.9)	16 (5.8)	--	277
Postgraduate diploma	256 (59.4)	137 (31.8)	37 (8.6)	1 (0.2)	431
Bachelor degree	937 (62.3)	457 (30.4)	108 (7.2)	2 (0.1)	1504
Undergraduate diploma	400 (61.5)	192 (29.5)	56 (8.6)	2 (0.3)	650
Associate diploma	447 (56.9)	253 (32.2)	86 (10.9)	--	786
Skilled/basic vocational	1790 (50.5)	1264 (35.7)	475 (13.4)	16 (0.5)	3545
ORIGIN					
Australia and New Zealand	11285 (53.8)	6948 (33.2)	2591 (12.4)	134 (0.6)	20958
British Isles and Ireland	1433 (54.4)	919 (34.9)	277 (10.5)	7 (0.3)	2636
Europe	995 (41.9)	981 (41.3)	384 (16.2)	14 (0.6)	2374
Middle East	89 (41.8)	90 (42.3)	33 (15.5)	1 (0.5)	213
Asia	1067 (76.4)	285 (20.4)	44 (3.1)	1 (0.1)	1397
Other	426 (53.4)	270 (33.8)	96 (12.0)	6 (0.8)	798
INCOME					
$30000-34999	1151 (52.7)	768 (35.2)	252 (11.5)	12 (0.5)	2183
NA, Don't know / Not Stated	1957 (53.5)	1268 (34.7)	422 (11.5)	12 (0.3)	3659
Negative	163 (48.5)	140 (41.7)	32 (9.5)	1 (0.3)	336
$1-4999	1108 (59.8)	497 (26.8)	235 (12.7)	12 (0.6)	1852
$5000-9999	2017 (51.1)	1259 (31.9)	622 (15.7)	53 (1.3)	3951
$10000-14999	1348 (58.3)	642 (27.8)	301 (13.0)	20 (0.9)	2311
$15000-19999	1335 (60.2)	642 (29.0)	226 (10.2)	14 (0.6)	2217
$20000-24999	1767 (56.5)	1018 (32.6)	329 (10.5)	11 (0.4)	3125
$25000-29999	1585 (55.8)	920 (32.4)	325 (11.4)	13 (0.5)	2843
$35000-39999	828 (51.9)	583 (36.5)	181 (11.3)	4 (0.3)	1596
$40000-44999	563 (47.9)	449 (38.2)	157 (13.4)	6 (0.5)	1175
$45000-49999	359 (49.9)	277 (38.5)	81 (11.3)	2 (0.3)	719
$50000-54999	303 (44.6)	290 (42.7)	83 (12.2)	3 (0.4)	679
$55000-59999	159 (47.0)	149 (44.1)	30 (8.9)	--	338
$60000-64999	160 (51.6)	110 (35.5)	40 (12.9)	--	310
$65000-69999	81 (46.0)	71 (40.3)	24 (13.6)	--	176
$70000-74999	82 (47.1)	76 (43.7)	16 (9.2)	--	174
$75000 or more	329 (44.9)	334 (45.6)	69 (9.4)	--	732

Table 5.1 (Cont.)

EMPLOYMENT STATUS					
Wage and salary	11576 (54.5)	7261 (34.2)	2323 (10.9)	84 (0.4)	21244
In own business	721 (55.9)	395 (30.6)	167 (12.9)	7 (0.5)	1290
Other/NA	2998 (51.3)	1837 (31.4)	935 (16.0)	72 (1.2)	5842
GEOGRAPHY					
Capital City	9221 (55.0)	5502 (32.8)	1952 (11.6)	99 (0.6)	16774
Large/small rural centres	1599 (51.7)	1042 (33.7)	436 (14.1)	17 (0.5)	3094
Other rural area/remote	1918 (49.3)	1416 (36.4)	532 (13.7)	28 (0.7)	3894
ACT/NT	2557 (55.4)	1533 (33.2)	505 (10.9)	19 (0.4)	4614
INSURANCE					
Does not have private insurance	4310 (54.8)	2521 (32.1)	975 (12.4)	54 (0.7)	7860
N/A	7545 (54.0)	4665 (33.4)	1671 (12.0)	84 (0.6)	13965
Has private insurance cover	3440 (52.5)	2307 (35.2)	779 (11.9)	25 (0.4)	6551
RISK FACTORS					
Diabetes Mellitus (Type 2)	51 (18.8)	109 (40.2)	97 (35.8)	14 (5.2)	271
Hypertension	802 (30.2)	1055 (39.8)	742 (28.0)	55 (2.1)	2654
High Cholesterol	305 (41.6)	292 (39.8)	129 (17.6)	7 (1.0)	733
Coronary Heart Disease	613 (34.6)	786 (44.3)	355 (20.0)	20 (1.1)	1774
Depression	274 (50.3)	174 (31.9)	90 (16.5)	7 (1.3)	545
Musculoskeletal disorders	2929 (44.4)	2479 (37.6)	1116 (16.9)	71 (1.1)	6595

Note: Sample characteristics include BMI ≥ 18.5 and Aged 20 to 64.

5.3 Logistic Model Estimates: The Relationship between RF and BMI

In total, 12 logistic equations were estimated to assess the relationship between RF and BMI. As noted in the previous chapter, the BMI coefficients are of particular interest in all these models (i.e., β_9 from equations 4.20 to 4.25 in chapter 4). As presented in Table 4.1, the principal hypothesis to test is whether BMI affects RF. The association between BMI and RF is summarised in Table 5.2 (complete details for *all* coefficients – including control variables – are provided in Tables A1 to A12 in Appendix A, on CD-ROM).

Table 5.2: Logistic regression estimates (β), standard errors (SE), and odds ratios (OR) (n = 28376); dependent variables = risk factors

	β	*SE*	*OR*	*Label*
Type 2 Diabetes Mellitus				
(Equation 4.20)				
BMI1	0.9211***	0.1741	2.5121	BMI 25 to < 30
BMI2	1.7530***	0.1791	5.7719	BMI 30 < 40
BMI3	2.9320***	0.3327	18.7660	BMI ≥ 40
#BMI***				
BMI	0.1480***	0.01 5	1.1595	Continuous measure
Hypertension				
(Equation 4.21)				
BMI1	0.5712***	0.05 9	1.7704	BMI 25 to < 30
BMI2	1.3650***	0.0593	3.9155	BMI 30 < 40
BMI3	2.0715***	0.1869	7.9367	BMI ≥ 40
#BMI***				
BMI	0.1193***	0.0048	1.1267	Continuous measure
CHD				
(Equation 4.22)				
BMI1	0.1184	0.0865	1.1257	BMI 25 to < 30
BMI2	0.2534**	0.1108	1.2883	BMI 30 < 40
BMI3	0.4459	0.4012	1.5619	BMI ≥ 40
#BMI*				
BMI	0.0231***	0.0089	1.0234	Continuous measure
High Cholesterol				
(Equation 4.23)				
BMI1	0.4683***	0.0583	1.5972	BMI 25 to < 30
BMI2	0.6968***	0.0728	2.0073	BMI 30 < 40
BMI3	1.0248***	0.2525	2.7865	BMI ≥ 40
#BMI***				
BMI	0.0623***	0.0057	1.0643	Continuous measure
Depression				
(Equation 4.24)				
BMI1	0.0400	0.1014	1.0409	BMI 25 to < 30
BMI2	0.2062	0.1257	1.2290	BMI 30 < 40
BMI3	0.2614	0.3935	1.2988	BMI ≥ 40
#BMI				
BMI	0.0163*	0.0095	1.0164	Continuous measure
Musculoskeletal Pain				
(Equation 4.25)				
BMI1	0.2605***	0.0333	1.2976	BMI 25 to < 30
BMI2	0.4799***	0.0443	1.6159	BMI 30 < 40
BMI3	0.8512***	0.1667	2.3425	BMI ≥ 40
#BMI***				
BMI	0.0452***	0.0035	1.0463	Continuous measure

#BMI – Joint hypothesis tests for the discrete BMI variables as a group.
* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$

Included in Table 5.2 are the logistic regression estimates (β), standard errors (SE), and odds ratios (OR) for the BMI variables. BMI was estimated as both a discrete and continuous measure. For the discrete measure of BMI, joint hypothesis tests were conducted to assess whether these variables, as a group, were statistically significant. In addition, each discrete BMI variable was compared to the excluded reference group (i.e., those individuals with a BMI value 18.5 to 24.9).

Overall, there is a statistically significant association between the level of BMI and the likelihood that an individual has an obesity-related risk factor (RF). For the discrete measure of BMI, these results clearly indicate that higher levels of BMI are associated with the following four risk factors: type 2 diabetes mellitus, hypertension, high cholesterol, and musculoskeletal pain (all p values are < 0.01). Moreover, the joint hypothesis tests for BMI, as a group, also indicate that increasing BMI is associated with these particular risk factors (p values < 0.01).

With respect to the discrete BMI variables, the odds ratios presented in Table 5.2, can be readily interpreted. For illustrative purposes, consider the BMI1 coefficient (and corresponding odds ratio) for type 2 diabetes mellitus (Equation 4.20). This result can be interpreted as a follows: compared to those individuals in a healthy weight range (i.e., BMI 18.5 to 24.9), overweight individuals (BMI1) are 2.5 times more likely to have type 2 diabetes mellitus. Appropriate interpretations can be made for all other discrete BMI coefficients so long as each coefficient is compared to the excluded reference group.

More importantly, higher levels of BMI are associated with a greater likelihood that an individual will have one of these particular risk factors. For instance, individuals in the outermost BMI group (i.e., those individuals with a BMI greater than or equal to 40) are much more likely to have a risk factor. Surprisingly, as a group, the three discrete BMI variables are not associated with coronary heart disease (CHD) at the 5 per cent level (although they are statistically significant at the 10 per cent level). However, it is interesting to note that when Equation (4.22) was estimated using BMI as a continuous variable this relationship was statistically significant ($p < 0.01$). This particular finding indicates that information is lost through the categorisation of the continuous BMI variable. Moreover, as a group, the three discrete BMI variables are not associated with depression but the continuous measure of BMI is associated with depression at the 10 per cent level of significance.

It is also worth noting that the logistic coefficient associated with the continuous measure of BMI can be interpreted as the change in the dependent variable, logistic (Y), associated with a one-unit change in the explanatory variable (Hamilton, 1992; Menard, 1995). For example, the continuous BMI coefficient in Table 5.2, for Equation (4.20), suggests that for a one-unit increase in BMI, the log of the odds in favour of having type 2 diabetes mellitus goes up by about 0.1480. Taking the antilog of 0.1480 results in an odds ratio of 1.1595. This means that for a unit increase in BMI, the odds in favour of developing type 2 diabetes mellitus increase by about 1.1595 or 15.95per cent. Similar interpretations can be made for all estimated continuous BMI coefficients presented in Table 5.2. Moreover, the estimated logistic coefficients can also be interpreted as probabilities. Therefore, it is possible to calculate the associated probabilities of having, for example, type 2 diabetes mellitus at different levels of BMI.

In other words, it is possible to map the probability of developing type 2 diabetes mellitus against body mass. For illustrative purposes, the estimated coefficients for Equation (4.20) using BMI as a continuous measure were selected (refer to Table A7 in Appendix A for complete details). As a group, only the age and employment status variables were statistically significant ($p < 0.01$). However, for descriptive purposes, all the information contained in Table A7 was used to estimate the probability of having type 2 diabetes conditional upon a set of personal characteristics and a given value of BMI. Taking this into account, the logistic Equation from Table A7 is approximately:

$$L_i = -11.79 + .51X_{i1} + .87X_{i2} + 1.22X_{i3} + 2.53X_{i4} + 2.66X_{i5} + 3.20X_{i6} + 3.11X_{i7} + 3.21X_{i8} + .40X_{i9} + .66X_{i10} + .01X_{i11} + .26X_{i12} + .09X_{i13} - .18X_{i14} - .28X_{i15} + .05X_{i16} - .004X_{i17} + .05X_{i18} - .89X_{i19} + .34X_{i20} + .64X_{i21} + .18X_{i22} - .53X_{i23} + .09X_{i24} + .50X_{i25} + .18X_{i26} + .05X_{i27} + .43X_{i28} - .18X_{i29} - .08X_{i30} - .39X_{i31} + .61X_{i32} - .08X_{i33} + .56X_{i34} - .78X_{i35} + .82X_{i36} - .21X_{i37} - .69X_{i38} + .29X_{i39} + .70X_{i40} - .07X_{i41} - .10X_{i42} + .11X_{i43} - .52X_{i44} - .11X_{i45} + .15X_{i46} \quad (5.0)$$

where variables are described as follows:

	Age
X_1	25-29 years
X_2	30-34 years
X_3	35-39 years
X_4	40-44 years
X_5	45-49 years
X_6	50-54 years
X_7	55-59 years
X_8	60-64 years
	Sex
X_9	Female
	Education
X_{10}	NA, inadequately described
X_{11}	Higher degree
X_{12}	Postgraduate diploma
X_{13}	Bachelor degree
X_{14}	Undergraduate diploma
X_{15}	Associate diploma
X_{16}	Skilled/basic vocational
	Origin
X_{17}	British Isles and Ireland
X_{18}	Europe
X_{19}	Middle East
X_{20}	Asia
X_{21}	Other
X_{22}	NA, Don't know/Not Stated
	Income
X_{23}	Negative
X_{24}	$1-4999
X_{25}	$5000-9999
X_{26}	$10000-14999
X_{27}	$15000-19999
X_{28}	$20000-24999
X_{29}	$25000-29999
X_{30}	$35000-39999
X_{31}	$40000-44999
X_{32}	$45000-49999
X_{33}	$50000-54999
X_{34}	$55000-59999
X_{35}	$60000-64999
X_{36}	$65000-69999
X_{37}	$70000-74999
X_{38}	$75000 or more
	Employment Status
X_{39}	In own business
X_{40}	Other/NA
	Geographic Location
X_{41}	Large/small rural centres
X_{42}	Other rural area/remote
X_{43}	ACT/NT
X_{44}	N/A
	Insurance Cover
X_{45}	Has private insurance cover
	BMI
X_{46}	Continuous measure of BMI, and
L_i	Predicted log odds of type 2 diabetes mellitus

The variables X_1-X_{45} are the discrete control variables that have been defined in the previous chapter. The variable X_{46} is the continuous measure of the BMI. Thus, Equation (5.0) specifies the predicted log odds of type 2 diabetes mellitus for a person, as a function of BMI and other factors (control variables). Consider the case of a male (X_9 =0) with a BMI of 30 (X_{46} = 30), and the following personal characteristics:

- age 50-54 (X_6 = 1 and X_1-X_5, X_7-X_8 = 0);
- no higher qualifications (X_{10}-X_{16} = 0);
- born in Australasia (X_{17}-X_{21} = 0);
- wage earner (X_{39}-X_{40} = 0);
- gross income per year $25,000-29,999 ($X_{29}$ = 1 and X_{22}-X_{28}, X_{30}-X_{38} = 0);
- residing in a large/small rural centre (X_{41} = 1 and X_{42}-X_{43} = 0); and
- no private health insurance (X_{44}-X_{45} = 0).

Using these characteristics, Equations (5.0) reduces to:

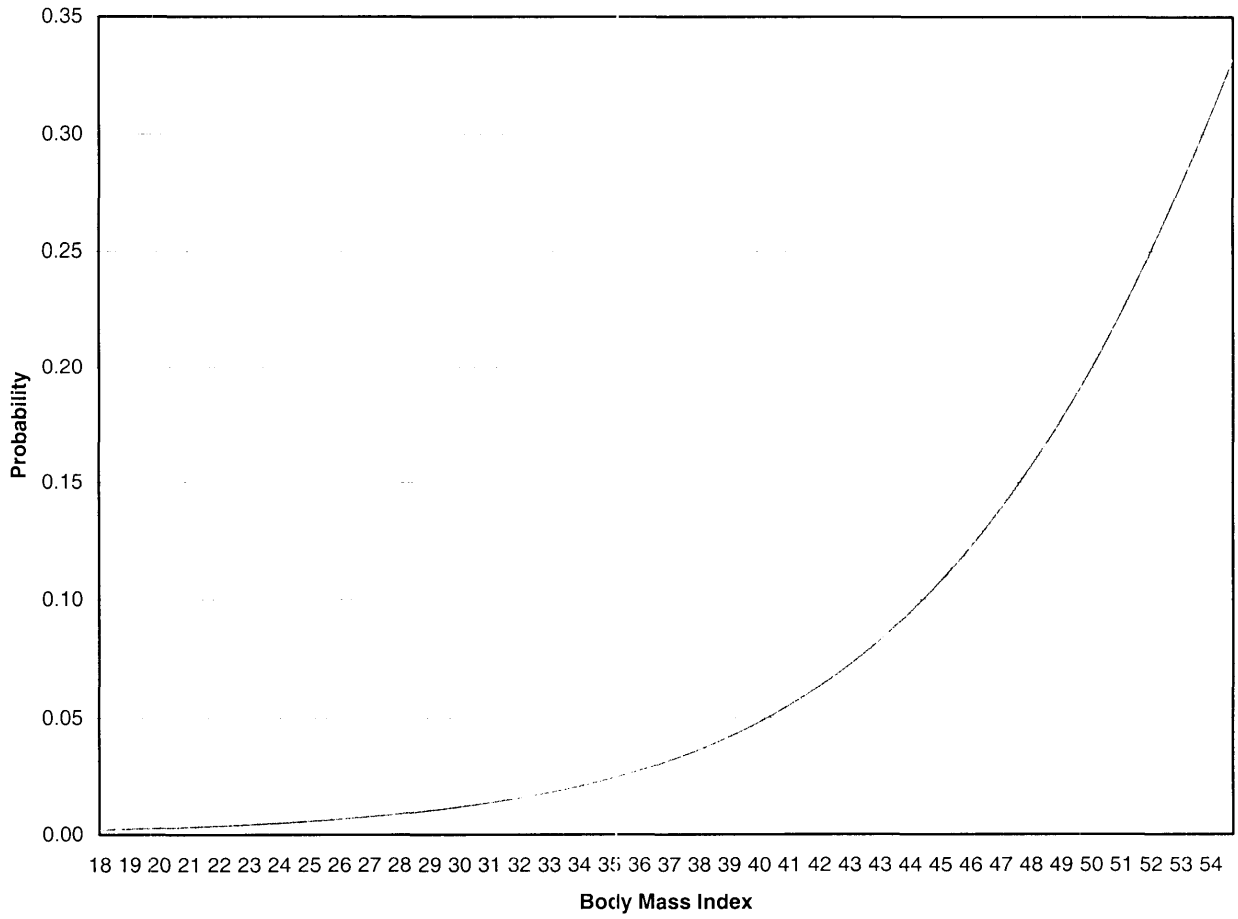
$$L_i = -11.79 + 3.20(1) - .18(1) - .07(1) + .15(30) = -4.34 \qquad (5.1)$$

The value of L_i in (5.1) gives the predicted log odds of type 2 diabetes (L_i) as -4.34. This can then be translated into the following probability: $e^{-4.34}/(1+e^{-4.34}) = 0.0121$. Thus, the probability that an individual with these characteristics and a BMI of 30 has type 2 diabetes is equal to 0.0121. The primary advantage of this approach is that it is relatively easy to interpret. This relationship between BMI and the probabilities for a person with these characteristics is summarised in Figure 5.1 below.

As shown in Figure 5.1, increasing BMI will lead to an increase in the probability that an individual with the above characteristics will develop type 2 diabetes mellitus. The increase in the probability over the BMI range of 20-40 is relatively small. However,

for those individuals, with a BMI greater than 40, the probability of developing type 2 diabetes increases significantly.

Figure 5.1: Probability of type 2 diabetes mellitus



For comparative purposes, Table 5.3 (below) provides the probabilities associated with having an obesity-related risk factor at different levels of BMI. For consistency, the above personal characteristics were also selected for each remaining risk factor. As shown in Table 5.3, the probabilities for each risk factor increase as the level of BMI increases (and at an increasing rate). Particularly striking relationships are found for hypertension, high cholesterol, and musculoskeletal pain.

Table 5.3: Probability of risk factors by level of BMI

Risk Factors	Level of BMI	Probability of having risk factor
Type 2 Diabetes Mellitus	20	0.0028
	25	0.0058
	30	0.0121
	35	0.0251
	40	0.0512
	45	0.1016
	50	0.1917
Hypertension	20	0.0838
	25	0.1424
	30	0.2316
	35	0.3537
	40	0.4984
	45	0.6434
	50	0.7661
CHD	20	0.0298
	25	0.0333
	30	0.0373
	35	0.0416
	40	0.0465
	45	0.0519
	50	0.0579
High Cholesterol	20	0.0558
	25	0.0747
	30	0.0992
	35	0.1308
	40	0.1704
	45	0.2190
	50	0.2769
Depression	20	0.0061
	25	0.0066
	30	0.0072
	35	0.0078
	40	0.0085
	45	0.0092
	50	0.0099
Musculoskeletal Pain	20	0.2616
	25	0.3075
	30	0.3576
	35	0.4110
	40	0.4666
	45	0.5231
	50	0.5789

Note: Probabilities are for men, age between 50-54, with no higher qualifications, born in Australasia, earning a wage between $25,000-29,999, residing in a large/small rural centre, with no private health insurance.

For example, the probability that an individual with these characteristics has hypertension at a BMI level of 30 is equal to 0.23. If this particular individual has a BMI value of 40 then the corresponding probability is almost 0.50. Similar interpretations can be made for all the probabilities presented in Table 5.3. Overall, these results strongly support the hypothesis that increasing BMI lead to an increasing probability of having each risk factor.

5.3.1 Diagnostic Statistics

It is useful to examine the summary measures for evaluating the logistic regression models. The principal concern is how well these models fit the data.

In logistic regression, there are close parallels to the statistics F and R^2 in linear regression. In linear regression, the sum of squared errors is the criterion used to select variables (Green 1993; Gujarati, 1995; Menard, 1995; Selvin, 1995). However, in the case of logistic regression, the log-likelihood (LL) statistic is used to select variables (Menard, 1995). Most statistical programs usually report not the log-likelihood statistic but the log-likelihood statistic multiplied by –2 (i.e., -2LL). This is because when the log-likelihood statistic is multiplied by –2 it has an approximate χ^2 (chi-square) distribution (Menard, 1995; Selvin, 1995).

The value of the –2LL statistic for the logistic regression model with only the intercept term included (hereafter referred to as D_0) is called the 'Initial Log Likelihood Function' in the SPSS logistic regression output. This statistic provides information when *none* of the explanatory variables are included in the regression equation. This

particular statistic is analogous to the total sum of squares (TSS) in linear regression (Menard, 1995).

The value of –2LL for the logistic model that includes all explanatory variables as well as the intercept term is called the '-2 Log Likelihood' in the SPSS output. For notational convenience, this particular statistic will be referred to as D_M. This statistic is comparable to the error sum of squares (SSE) in linear regression (Menard, 1995). In short, D_M is used as an indicator of how poorly the model fits with all the explanatory variables included in the logistic equation (Menard, 1995).

In logistic regression, the difference between D_0 and D_M (i.e., $D_0 - D_M$) is called the 'Model Chi-Square' in the SPSS output (Menard, 1995). From this point on, the difference will be referred to as G_M. This particular statistic is comparable to the *F* test in linear regression (Selvin, 1995). Specifically, G_M tests the null hypothesis that $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ for the logistic regression model (Menard, 1995; Selvin, 1995). Therefore if G_M is statistically significant at the 5 per cent level, one can reject the null hypothesis and conclude that the information about the explanatory variables allows one to make a better prediction than one could *without* the explanatory variables. According to Menard (1995) it is "advisable to focus primarily on G_M and only secondarily on D_M" (Menard, 1995, p.21).

It is also worth noting that several analogues to the R^2 measure of goodness-of-fit in linear regression have been proposed for logistic regression (Green, 1993; Menard, 1995). However, to maintain the current analogy between linear and logistic regression, the R^2_L proposed by Hosmer and Lemeshow (1989) provides one measure of the

association between the independent and explanatory variables in logistic regression. Formally, $R^2_L = G_M/(D_0) = G_M/(G_M+D_M)$. Menard (1995) states that the:

> "R^2_L is the proportional reduction in χ^2 or a proportional reduction in the absolute value of the log-likelihood measure. It indicates by how much the inclusion of the independent variables in the model reduces the badness-of-fit D_0 chi-square statistic" (Menard, 1995, p.22).

Therefore, the above summary statistics can be used to examine how well the models fit the relationship between RF and BMI. For convenience, this information is summarised in Table 5.4 below.

Table 5.4: Logistic diagnostic statistics for the relationships between RF and BMI

Equation	D_0	D_M	G_M	R^2_L
Type 2 diabetes mellitus (Equation 4.20)				
BMI as discrete measure	3060.34	2545.07	515.27	0.1688
BMI as continuous measure	3060.34	2530.11	530.23	0.1733
Hypertension (Equation 4.21)				
BMI as discrete measure	17628.83	14693.42	2935.41	0.1665
BMI as continuous measure	17628.83	14657.52	2971.31	0.1685
CHD (Equation 4.22)				
BMI as discrete measure	6806.82	6088.17	718.65	0.1056
BMI as continuous measure	6806.82	6087.77	719.06	0.1056
High Cholesterol (Equation 4.23)				
BMI as discrete measure	13270.85	11746.95	1523.90	0.1148
BMI as continuous measure	13270.85	11751.88	1518.97	0.1145
Depression (Equation 4.24)				
BMI as discrete measure	5387.70	5020.94	366.77	0.0681
BMI as continuous measure	5387.70	5020.89	366.82	0.0681
Musculoskeletal Pain (Equation 4.25)				
BMI as discrete measure	30769.67	28450.80	2318.87	0.0754
BMI as continuous measure	30769.67	28433.72	2335.95	0.0759

Notes: D_0 = The value of the –2LL statistic for the logistic regression model with only the intercept term included. D_M = The value of –2LL for the logistic model that includes all explanatory variables as well as the intercept term. G_M = The difference between D_0 and D_M. $R^2_L = G_M/(D_0) = G_M/(G_M+D_M)$. All the GM statistics are significant at the 1 per cent level.

Included in Table 5.4 are the D_0, D_M, G_M, and R^2_L logistic regression diagnostics. These statistics are presented for logistic equations estimated with body mass index

(BMI) as a discrete and continuous measure. All G_M statistics are significant at the 1 per cent level (i.e., $p < 0.01$). Therefore, one can reject the null hypothesis and conclude that the information about the control and BMI variables allows one to make a better prediction than one could *without* these variables. The R^2_L diagnostics indicate that models (4.20) to (4.23) fit the data reasonably well, especially considering that the data are cross sectional survey data. Overall, these measures provide information about how well the logistic regression models actually fit the data.

5.4 Logistic Model Estimates: The Relationship between MCU and BMI

Logistic equations were also estimated for the three different types of medical care utilisation (MCU), namely doctor visits, other health care professional visits, and hospital visits. As already noted, the BMI and RF coefficients are of particular interest in these logistic models (i.e, β_9 and β_{10} from equations 4.27 to 4.29 in chapter 4). As indicated in Table 4.1, the hypotheses tested are: (1) whether BMI affects RF, (2) whether RF affects MCU, and (3) whether BMI has any direct effect on MCU. Results for each type of MCU are presented below. The Tables presented in this section summarise the relationship between MCU, BMI, and RF (Complete details for all coefficients are reported in Appendix B, which is contained on CD-ROM attached to the inside of the back cover).

5.4.1 Logistic Model Estimates for Doctor Visit

Logistic regression estimates for the relationship between the likelihood of a doctor visit and BMI are presented in Table 5.5. It should be noted that Equation (4.27) was initially estimated with: (i) control and discrete BMI variables only; (ii) control and RF variables only; and (iii) control, RF, and discrete BMI variables. For convenience, these

alternative models have been labelled (1), (2), and (3), respectively. For each model, the parameter estimates (β) and standard errors (SE) are reported. Maximum likelihood (ML) ratio test results for model comparison are also presented. The ML ratio tests the significance of either including or removing a group of variables. Finally, the R^2_L for each model is also reported. Note that the coefficients in Table 5.5 are shown for *only* the RF and BMI variables. (Complete details for *all* variables – including the control variables – are located in Appendix B).

Table 5.5: Logistic regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = doctor visit (with discrete BMI variables)

	Model (1)		Model (2)		Model (3)	
	β	**SE**	β	**SE**	β	**SE**
Doctor Visit (Equation 4.27)						
BMI Variables						
BMI1	0.1177***	0.0334			0.0620*	0.0340
BMI2	0.2207***	0.0454			0.0655	0.0472
BMI3	0.4559***	0.1680			0.1539	0.1748
Risk Variables						
DM_TY2			0.6122***	0.1325	0.5947***	0.1331
HYPER			0.4840***	0.0488	0.4715***	0.0494
DEPRESS			1.0805***	0.0904	1.0801***	0.0905
CHD			0.5977***	0.0826	0.5992***	0.0826
CHOLEST			0.3183***	0.0585	0.3127***	0.0586
MS_DIS			0.5276***	0.0341	0.5225***	0.0342
LL	-14567.84		-14252.36		-14250.1	
DF	48		51		54	
ML Ratio Tests	--		--		(1) vs. (3)	$p < 0.01$
	--		--		(2) vs. (3)	NS
R^2_L	0.0268		0.0478		0.0480	

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood. NS = not statistically significant.

The estimates for model 1 in Table 5.5 indicate that there is a statistically significant relationship between the likelihood of a doctor visit in 1995 and the level of BMI. This

association is particularly evident for individuals in the upper two BMI categories (i.e., variables BMI2 and BMI3). Individuals within the BMI range of 30-39.9 (BMI2) were 25 per cent (OR = 1.25) more likely to have visited a doctor in 1995, compared to individuals within a healthy BMI range (18.5-24.9). Individuals in the outermost BMI group (BMI3) were 58 per cent (OR = 1.58) more likely to have had a doctor visit in 1995, compared to individuals within the healthy BMI range. For model 1, the corresponding goodness of fit measure, R^2_L, is 0.0268.

Model 2, which includes only control and RF variables, clearly indicates that there is a strong relationship between the likelihood of a doctor visit and risk factors. All risk factor variables in this model are statistically significant at the 1 per cent level. The benefit of estimating model 2 is that it can be compared to model 3 to ascertain whether the BMI variables, as a group, are statistically significant. The R^2_L for model 2 is 0.0478.

Model 3, which includes the control, BMI, and RF variables, indicates that all the RF variables are statistically significant at the 1 per cent level. While the three discrete BMI variables are not statistically significant at the 5 per cent level, the BMI1 variable is statistically significant at the 10 per cent level. The corresponding ML ratio test comparing models 2 and 3 is not statistically significant (that is, testing whether the BMI variables as a group are statistically significant). There is no evidence of an independent BMI effect.

The results presented in Table 5.5 will now be discussed with reference to the working hypotheses presented in Table 4.1. To begin with, the estimation of model 3 tests

hypotheses 2 and 3 from chapter 4. Hypothesis 2 states that the presence of RF would lead to increased MCU. In other words, these risk factors affect MCU. Thus, the estimated coefficients on the RF variables should be positive. Since the coefficients on the RF variables in model 3 are positive, hypothesis 2 is not rejected. Therefore, one may conclude that there is an association between MCU and RF.

Hypothesis 3 states that BMI has no independent effect on MCU. Thus the estimated coefficients on the BMI variables in model 3 should not significantly differ from zero. Hypothesis 3 is not rejected (as supported by the ML ratio test) and it can be inferred that BMI have no independent impact on MCU.

Together, models 1 and 3 in Table 5.5 can be used to test hypothesis 1 from chapter 4. Hypothesis 1 states that BMI affects RF. As shown in Table 5.5, the BMI coefficients in model 1 differ from the BMI coefficients in model 3, indicating that BMI and RF are, in fact, correlated. In chapter 4 (section 4.2.3), it was demonstrated that if the BMI coefficients in model 1 differ from zero and the BMI coefficients in model 3 are relatively close to zero, then one can infer that BMI affects RF. Alternatively, in path model terms, if the introduction of the RF variables is statistically significant (model 3) and the BMI variables are attenuated, then this indicates that the impact of obesity is being exerted through these risk factors. Thus, the results presented in Table 5.5 provide evidence that there is causal relationship between BMI and RF. These findings support hypothesis 1 from Table 4.1 that BMI affects RF.

The relationship between the likelihood of a doctor visit and BMI was also estimated using BMI as a continuous variable. By using a continuous measure of BMI,

information is not lost through categorisation (i.e., transforming a continuous variable into a series of discrete variables). A continuous measure of BMI can also be used to isolate the marginal effect associated with a one-unit increase in BMI.

Table 5.6 below contains the logistic regression estimates for the relationship between the likelihood of doctor visit and BMI, with BMI estimated as a continuous measure. Table 5.6 is in the same format as Table 5.5. For each model, the parameter estimates (β) and standard errors (SE) are reported. Maximum likelihood (ML) ratio test results for model comparison are also presented. The R^2_L for each model is also reported.

The estimates presented in Table 5.6, indicate that there is a statistically significant positive association between the likelihood of a doctor visit and the level of BMI (model 1). The coefficient for the continuous BMI variable is 0.0198 ($p < 0.01$). This result suggests that for a one-unit increase in BMI, the log of the odds in favour of visiting a doctor goes up by about 0.0198. Taking the antilog of 0.0198 results in an odds ratio of 1.02. Thus, for a one-unit increase in BMI the odds in favour of visiting a doctor increase by about 1.02 or 2 per cent.

As shown in Table 5.6, all the RF variables were statistically significant at the 1 per cent level (models 2 and 3). In fact, the introduction of the RF variables (model 3) substantially attenuated the impact of the BMI variable. Although the corresponding BMI coefficient in model 3 is close to zero it is nevertheless statistically significant at the 10 per cent level. However, the corresponding ML ratio test comparing models 2 and 3 indicates that there is no evidence of an independent BMI effect. Table 5.6

clearly indicates that most of the impact of obesity is being exerted through these six risk factors.

Table 5.6: Logistic regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = doctor visit (with BMI as a continuous variable)

	Model (1)		Model (2)		Model (3)	
	β	**SE**	β	**SE**	β	**SE**
Doctor Visit (Equation 4.27)						
BMI Variable						
BMI	0.0198***	0.0035			0.0064*	0.0036
Risk Variables						
DM_TY2			0.6122***	0.1325	0.5930***	0.1330
HYPER			0.4840***	0.0488	0.4705***	0.0494
DEPRESS			1.0805***	0.0904	1.0795***	0.0905
CHD			0.5977***	0.0826	0.5992***	0.0826
CHOLEST			0.3183***	0.0585	0.3138***	0.0586
MS_DIS			0.5276***	0.0341	0.5227***	0.0342
LL	-14568.12		-14252.36		-14250.8	
DF	46		51		52	
ML Ratio Test	--		--		(1) vs. (3)	$p < 0.01$
	--		--		(2) vs. (3)	NS
R^2_L	0.0267		0.0478		0.0480	

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood. NS = not statistically significant.

5.4.2 Logistic Model Estimates for Health Care Professional Visit

Logistic regression estimates for the relationship between the likelihood of a health care professional visit (OHP) and BMI are presented in Table 5.7 below. Equation (4.28) was initially estimated with: (i) control and discrete BMI variables only; (ii) control and RF variables only, and (iii) control, RF, and discrete BMI variables. For each model, the parameter estimates (β) and standard errors (SE) are reported. Maximum likelihood (ML) ratio test results and the R^2_L is also reported.

Table 5.7: Logistic regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = other health care professional visit (with BMI as discrete variables)

	(1)		(2)		(3)	
	β	SE	β	SE	β	SE
OHP Visit (Equation 4.28)						
BMI Variables						
BMI1	0.0422	0.0443			-0.0146	0.0451
BMI2	0.1811***	0.0596			0.0458	0.0619
BMI3	-0.0205	0.2472			-0.3309	0.2544
Risk Variables						
DM_TY2			0.6647***	0.1659	0.6748***	0.1666
HYPER			0.1417**	0.0686	0.1404**	0.0695
DEPRESS			0.9589***	0.1039	0.9590***	0.1039
CHD			0.0797	0.1207	0.0792	0.1207
CHOLEST			0.0546	0.0823	0.0537	0.0824
MS_DIS			0.9420***	0.0426	0.9424***	0.0427
LL	-9590.69		-9302.06		-9030.67	
DF	48		51		54	
ML Ratio Test	--		--		(1) vs. (3)	p < 0.01
	--		--		(2) vs. (3)	NS
R^2_L	0.0155		0.0451		0.0453	

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood. NS = not statistically significant.

The results for model 1 in Table 5.7 show that only one discrete BMI variable, BMI2, is statistically significant ($p < 0.01$). Although this finding is somewhat surprising, it could suggest that the use of these discrete BMI variables (and subsequent loss of information through categorisation) may, in fact, be obscuring the underlying relationship between the likelihood of a health care professional visit and the level of BMI.

The estimation of models 2 and 3 indicate that the following RF variables are statistically significant at the 5 per cent level or better: type 2 diabetes mellitus,

hypertension, depression, and musculoskeletal disorders. The CHD and high cholesterol variables are not statistically significant. This finding may be due, in part, to the nature of these particular risk factors. It could be argued that individuals with these particular risk factors may not be inclined to visit other health care professionals. However, the ML ratio test comparing models 1 and 3 indicates that, as a group, the RF variables are statistically significant.

Furthermore, the BMI coefficients in model 3 are not statistically significant at the 5 per cent level. The ML ratio test comparing models 2 and 3 indicates that, as a group, the BMI variables are not statistically significant, indicating a failure to reject the null hypothesis of zero BMI effect.

While the estimation of models 1 and 3 to some extent support the working hypotheses presented in chapter 4, these results are less convincing. Subsequently, it was discovered that model improvement was obtained when Equation (4.28) was estimated using BMI as a continuous variable. These results are presented in Table 5.8 below. Table 5.8 has a similar format to Table 5.7, with appropriate interpretation.

Model 1 in Table 5.8 clearly indicates that the continuous BMI coefficient ($\beta = 0.015$) is statistically significant ($p < 0.01$). Taking the antilog of the BMI coefficient results in an odds ratio of 1.015. Therefore, for a one-unit increase in BMI the odds in favour of visiting a health care professional increase by about 1.015 or 1.5 per cent.

Table 5.8: Logistic regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = other health care professional visit (with BMI as a continuous variable)

	(1)		(2)		(3)	
	β	**SE**	β	**SE**	β	**SE**
OHP Visit (Equation 4.28)						
BMI Variable						
BMI	0.0150***	0.0045			0.0028	0.0047
Risk Variables						
DM_TY2			0.6647***	0.1659	0.6554***	0.1667
HYPER			0.1417**	0.0686	0.1352*	0.0695
DEPRESS			0.9589***	0.1039	0.9586***	0.1039
CHD			0.0797	0.1207	0.0807	0.1207
CHOLEST			0.0545	0.0823	0.0522	0.0824
MS_DIS			0.9423***	0.0426	0.9400***	0.0427
LL	-9589.9		-9302.06		-9301.5	
DF	46		51		52	
ML Ratio Test	--		--		(1) vs. (3)	$p < 0.01$
	--		--		(2) vs. (3)	NS
R^2_L	0.0156		0.0451		0.0452	

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood. NS = not statistically significant.

Model 3 in Table 5.8 shows that the following risk factors are statistically significant at the 1 per cent level: type 2 diabetes mellitus, depression, and musculoskeletal disorders. The hypertension variable is statistically significant but only at the 10 per cent level of significance ($p = 0.052$). The BMI coefficient in model 3 is not statistically significant and is relatively close to zero. The results for model 3 demonstrate that RF affects MCU and there is no evidence of an independent BMI effect. In fact, the inclusion of the RF variables attenuated the impact of the BMI variable. In path model terms, the introduction of the RF variables suggests that the impact of obesity is being exerted through these risk factors. This can be seen by comparing the BMI coefficients in models 1 and 3, respectively.

5.4.3 Logistic Model Estimates for Hospital Visits

Logistic regression estimates for the relationship between the likelihood of a hospital visit and the level of BMI are presented in Table 5.9. As with previous relationships presented in the above sections, Equation (4.29) was estimated with: (i) control and discrete BMI variables only; (ii) control and RF variables only, and (iii) control, RF, and discrete BMI variables. For each model the coefficients (β) and standard errors (SE) are reported. The ML ratio test results and the R^2_L for each model is also reported.

Table 5.9: Logistic regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = hospital visit (with BMI as discrete variables)

	(1)		(2)		(3)	
	β	**SE**	β	**SE**	β	**SE**
Hospital Visit (Equation 4.29)						
BMI Variables						
BMI1	0.0847	0.1618			0.0605	0.1627
BMI2	0.4112**	0.1955			0.3568*	0.1998
BMI3	0.2444	0.7243			0.1247	0.7330
Risk Variables						
DM_TY2			0.2744	0.5220	0.2154	0.5249
HYPER			0.1921	0.2199	0.1378	0.2222
DEPRESS			0.8566***	0.2972	0.8575***	0.2972
CHD			1.0884***	0.2688	1.0956***	0.2686
CHOLEST			-0.4318	0.3072	-0.4469	0.3070
MS_DIS			0.3078*	0.1589	0.2890*	0.1597
LL	-1201.3		-1189.98		-1188.44	
DF	48		51		54	
ML Ratio Test	--		--		(1) vs. (3)	$p < 0.01$
	--		--		(2) vs. (3)	NS
R^2_L	0.0459		0.0549		0.0561	

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$

Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood. NS = not statistically significant.

The results for model 1 as shown in Table 5.9 indicate that only one discrete BMI variable, namely BMI2, is statistically significant ($p < 0.05$). This finding suggests that the relationship between the likelihood of a hospital visits and the level of BMI is relatively weak. Model 3 in Table 5.9 shows that the depression and CHD variables are statistically significant at the 1 per cent level. The musculoskeletal disorder variable is statistically significant but only at the 10 per cent level ($p = 0.07$). No other risk factor variables are statistically significant. However, the ML ratio test (model 1 vs. model 3) clearly indicates that the introduction of the risk factor variables, as a group, is statistically significant ($p < 0.01$). Overall, the results in Table 5.9 indicate that there is a relationship between the likelihood of hospital visit and presence of risk factors (hypothesis 2) but the relationship is fairly weak. While model 3 demonstrates the BMI has no independent impact on MCU (hypothesis 3), there is little statistical evidence supporting hypothesis 1 that BMI affects RF.

With this in mind, Equation (4.29) was also estimated using BMI as a continuous variable. The results are reported in Table 5.10 below. Table 5.10 follows a similar format to Table 5.9 but appropriate interpretation is needed. The use of the continuous BMI variable makes little difference to the overall findings.

Model 1 in Table 5.10 clearly shows that the level of BMI is not associated with the likelihood of a hospital visit. However, the introduction of the RF variables, as a group, is statistically significant at the 1 per cent level as indicated by the corresponding ML ratio test (i.e., model 1 vs. model 3).

Table 5.10: Logistic regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = hospital visit (with BMI as a continuous variable)

	(1)		(2)		(3)	
	β	**SE**	β	**SE**	β	**SE**
Hospital Visit (Equation 4.29)						
BMI Variable						
BMI	0.0191	0.0155			0.0135	0.0159
Risk Variables						
DM_TY2			0.2744	0.5220	0.2322	0.5244
HYPER			0.1921	0.2199	0.1637	0.2223
DEPRESS			0.8566***	0.2972	0.8541***	0.2973
CHD			1.0884***	0.2688	1.0931***	0.2687
CHOLEST			-0.4313	0.3072	-0.4405	0.3072
MS_DIS			0.3078*	0.1589	0.2966*	0.1596
LL	2405.32		2379.96		2379.24	
DF	46		51		52	
ML Ratio Test	--		--		(1) vs. (3)	$p < 0.01$
	--		--		(2) vs. (3)	NS
R^2_L	0.0448		0.0549		0.0552	

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood. NS = not statistically significant.

Model 3 in Table 5.10 indicates that while the RF variables affect MCU there is no evidence to suggest that BMI has an independent impact on MCU. Surprisingly, there is no statistically positive relationship between BMI and the likelihood of a hospital visit (model 1). A possible explanation for this finding may be the limited number of dependent observations (recall that only 214 individuals were recorded as having spent at least one night in hospital during the survey period). Therefore, it is entirely possible that a larger sample may, in fact, uncover a statistically significant relationship between the likelihood of a hospital visit and the level of BMI (model 1).

5.5 Poisson Model Estimates: The Relationship between MCU and BMI

Initially, the level (or amount) of MCU was modelled using Poisson regression (Complete details for all Poisson coefficients are reported in Appendix C, on the CD-ROM.) Following this, censored Poisson regression techniques were also used to take into account the fact that dependent MCU variables were right-censored. The results for both the Poisson and censored Poisson estimates were similar. This is not surprising, as there were only a relatively small number of observations at the point of censoring (refer to Table 4.11). Thus, to avoid the repetition of discussing similar results, only the uncensored Poisson regression estimates are reported. (Complete details for all censored Poisson coefficients are presented in Appendix D, on the CD-ROM.)

Extensions to the Poisson technique were also examined (refer to Table 4.16 for a complete list of alternative models). However, it was discovered that little, if any, improvement was obtained by modifying the basic underlying Poisson model. Another point worth noting is that for the relationship between the number of hospital visits and BMI, i.e., equations (4.38) and (4.38a), the discrete income variables were re-coded owing to estimation problems. Subsequently, the 18 income categories were collapsed into 15 categories. As a result of this re-classification, 14 discrete income variables were used as opposed to 17. Details of the re-coded income variable are presented in the appropriate Tables contained within Appendices C and D. The Poisson estimates will now be discussed for each specific type of medical care utilisation.

5.5.1 Poisson Model Estimates for the Number of Doctor Visits

Poisson regression estimates for the relationship between the expected (mean) number of doctor visits and BMI are presented in Table 5.11. Equation (4.36) was estimated with: (i) control and discrete BMI variables only; (ii) control and RF variables only, and (iii) control, RF, and discrete BMI variables. For consistency, these alternative models have also been labelled (1), (2), and (3), respectively. For each model, the parameter estimates (β) and standard errors (SE) are reported. Maximum likelihood (ML) ratio test results for model comparison are also presented.

Table 5.11 : Poisson regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = number of doctor visits (with BMI as discrete variables)

	(1)		(2)		(3)	
	β	SE	β	SE	β	SE
Doctor Visits (Equation 4.36)						
BMI Variables						
BMI1	0.0795***	0.0251			0.0299	0.0253
BMI2	0.1776***	0.0331			0.0544	0.0338
BMI3	0.3160***	0.1121			0.0751	0.1133
Risk Variables						
DM_TY2			0.2970***	0.0832	0.2848***	0.0838
HYPER			0.3051***	0.0338	0.2957***	0.0343
DEPRESS			0.8090***	0.0487	0.8082***	0.0487
CHD			0.4078***	0.0520	0.4094***	0.0520
CHOLEST			0.2469***	0.0395	0.2433***	0.0396
MS_DIS			0.4424***	0.0244	0.4388***	0.0245
LL	-20131.62		-19719.59		-19717.5	
DF	48		51		54	
ML Ratio Test	--		--		(1) vs. (3)	p < 0.01
	--		--		(2) vs. (3)	NS

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood. NS = not statistically significant.

As presented in Table 5.11, Model 1 clearly indicates that there is a statistically significant positive relationship between the expected (or mean) number of doctor visits and the level of BMI (all BMI coefficients are statistically significant at the 1 per cent level).

Model 2, which includes only control and RF variables, also indicates that there is a strong relationship between the expected number of doctor visits and obesity-related risk factors. All RF variables in model 2 are statistically significant the 1 per cent level.

Model 3, which includes the control, BMI, and RF variables, indicates that all the RF variables are statistically significant the 1 per cent level. However, as expected, the three discrete BMI variables are not statistically significant. The corresponding ML ratio test comparing models 2 and 3 is not statistically significant. This indicates that the BMI variables, as a group, are not statistically significant. Furthermore, the BMI coefficients are relatively close to zero. The results for model 3 demonstrate that RF affects MCU and that BMI has little independent impact on MCU. In fact, the inclusion of the RF variables attenuated the impact of the BMI variables (this can be seen by comparing models 1 and 3). In path model terms, the introduction of the RF variables suggests that the impact of obesity is exerted through these particular risk factors (model 3).

Equation (4.36) was also estimated using BMI as a continuous measure. The results for all three models are presented in Table 5.12 below. Table 5.12 is in the same format as Table 5.11 above but appropriate interpretation is required. The use of the continuous BMI variable makes no real difference to the overall results.

Model 1, in Table 5.12, clearly indicates that there is a statistically significant relationship between the expected number of doctor visits and the continuous measure of BMI ($p < 0.01$). The corresponding BMI coefficient is 0.0153. This result can be interpreted as follows: a one-unit increase in BMI will result in a 1.5 per cent increase in the expected number of doctor visits.

Table 5.12: Poisson regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = number of doctor visits (with BMI as a continuous variable)

	(1)		(2)		(3)	
	β	**SE**	β	**SE**	β	**SE**
Doctor Visits (Equation 4.36)						
BMI Variable						
BMI	0.0153***	0.0025			0.0044*	0.0026
Risk Variables						
DM_TY2			0.2970***	0.0832	0.2828***	0.0836
HYPER			0.3051***	0.0338	0.2955***	0.0342
DEPRESS			0.8090***	0.0487	0.8079***	0.0487
CHD			0.4078***	0.0520	0.4096***	0.0520
CHOLEST			0.2469***	0.0395	0.2435***	0.0396
MS_DIS			0.4424***	0.0244	0.4387***	0.0245
LL	-20131.06		-19719.59		-19718.1	
DF	46		51		52	
ML Ratio Test	--		--		(1) vs. (3)	$p < 0.01$
	--		--		(2) vs. (3)	NS

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood. NS = not statistically significant.

In model 3 all RF variables are statistically significant at the 1 per cent level. The continuous BMI variable is statistically significant at the 10 per cent level. However, when comparing models 1 and 3 it is evident the that the introduction of the RF variables has attenuated the impact of BMI on MCU. Once again, this is consistent with the proposition that the damaging impact of obesity is exerted through the obesity-

associated risk factors. In addition, the corresponding ML ratio test comparing models 2 and 3 is not statistically significant, indicating that there is no evidence of an independent BMI effect.

5.5.2 Poisson Model Estimates for the Number of Health Care Professional Visits

Poisson regression estimates for the relationship between the expected number of other health care professional visits (OHP) and the level of BMI are reported in Table 5.13 below. For consistency, Equation (4.37) was estimated with: (i) control and discrete BMI variables only; (ii) control and RF variables only, and (iii) control, RF, and discrete BMI variables. Once again, the BMI and RF coefficients are of particular interest.

As reported in Table 5.13, model 1 shows that the expected number of health care professional visits is associated with only two discrete BMI variables, namely BMI1 and BMI2 ($p < 0.01$). Models 2 and 3 indicate that the following risk factor variables are statistically significant at the 5 per cent level or better: type 2 diabetes mellitus, hypertension, depression, and musculoskeletal disorders. The CHD and cholesterol risk factor variables are not statistically significant.

The results for model 3 are of particular interest and there are several points worth noting. First, the introduction of the RF variables, as a group, is statistically significant as indicated by the corresponding ML ratio test ($p< 0.01$). This result provides evidence that the RF variables impact upon the level of MCU. Furthermore, the inclusion of the RF variables also attenuates the impact of the BMI coefficients in model 3. Once again, this is consistent with previous findings, demonstrating that the impact of obesity is

partly operating through these particular risk factors. However, as a group, the discrete BMI variables in model 3 are statistically significant as shown by the ML ratio test comparing models 2 and 3 ($p < 0.025$). Thus while the impact of BMI is attenuated with the inclusion of the RF variables, there is evidence to suggest that obesity in and of itself exerts an independent impact on MCU. This is an interesting result and suggests that obese individuals may actually visit a health care professional in an attempt to specifically treat their weight.

Table 5.13: Poisson regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = number of other health care professional visits (with BMI as discrete variables)

	(1)		(2)		(3)	
	β	**SE**	β	**SE**	β	**SE**
OHP Visits (Equation 4.37)						
BMI Variables						
BMI1	0.1041***	0.0332			0.0495	0.0334
BMI2	0.2381***	0.0439			0.1076**	0.0449
BMI3	-0.0580	0.1907			-0.3148	0.1922
Risk Variables						
DM_TY2			0.2613**	0.1299	0.2661**	0.1306
HYPER			0.1715***	0.0498	0.1587***	0.0504
DEPRESS			0.8127***	0.0666	0.8129***	0.0666
CHD			-0.0517	0.0915	-0.0505	0.0915
CHOLEST			-0.0005	0.0613	-0.0081	0.0613
MS_DIS			0.9630***	0.0307	0.9590***	0.0308
LL	-14977.7		-14442.29		-14437.21	
DF	48		51		54	
ML Ratio Test	--		--		(1) vs. (3)	p< 0.01
	--		--		(2) vs. (3)	p < 0.025

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood.

It is of interest to note that the BMI3 coefficients in models 1 and 3 are not statistically significant and are negative. Once more, this may suggest that using a discrete measure of body mass may, in fact, be obscuring the underlying relationship. With this in mind,

Equation (4.37) was re-estimated using a continuous measure of BMI. The results for all three models are presented in Table 5.14.

Model 1, in Table 5.14, clearly shows that there is a positive statistical relationship between the expected number of visits to health care professionals and BMI ($p < 0.01$). In this instance, the use of a continuous measure of BMI is preferable to a discrete measurement. The corresponding BMI coefficient is 0.0212. This result can be interpreted as follows: a one-unit increase in BMI (holding all other explanatory variables fixed) will lead to a 2 per cent increase in the expected number of visits to health care professionals.

Table 5.14: Poisson regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = number of other health care professional visits (with BMI as a continuous variable)

	(1)		(2)		(3)	
	β	**SE**	β	**SE**	β	**SE**
OHP Visits (Equation 4.37)						
BMI Variable						
BMI	0.0212***	0.0033			0.0098***	0.0034
Risk Variables						
DM_TY2			0.2613**	0.1299	0.2241*	0.1306
HYPER			0.1715***	0.0498	0.1476***	0.0504
DEPRESS			0.8127***	0.0666	0.8117***	0.0666
CHD			-0.0517	0.0915	-0.0465	0.0915
CHOLEST			-0.0005	0.0613	-0.0096	0.0613
MS_DIS			0.9630***	0.0307	0.9559***	0.0308
LL	-14973.47		-14442.29		-14438.18	
DF	46		51		52	
ML Ratio Test	--		--		(1) vs. (3)	$p < 0.01$
	--		--		(2) vs. (3)	$p < 0.01$

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood.

Model 3 in Table 5.14 shows that the following RF variables are statistically significant: type 2 diabetes mellitus, hypertension, depression, and musculoskeletal pain. Moreover, as a group, the introduction of the RF variables is statistically significant as demonstrated by the ML ratio test comparing model 1 to model 3 ($p < 0.01$). The inclusion of the RF variables attenuates the impact of BMI on the expected number of visits to health care professionals even though the BMI coefficient in model 3 is statistically significant at the 1 per cent level. The ML ratio test comparing models 2 to 3 ($p < 0.021$) indicates that there is an independent BMI effect.

5.5.3 Poisson Model Estimates for the Number of Hospital Visits

Poisson regression estimates for the relationship between the expected number of hospital visits and the level of BMI (Equation 4.38) are presented in Table 5.15 below.

Table 5.15, is presented in the same format as previous tables but appropriate interpretation is required. To begin with, model 1 in Table 5.15 indicates that only two discrete BMI variables are statistically significant at the 1 per cent level. Interestingly, while the BMI coefficient in model 1 is statistically significant it is also negative, which is unexpected. Again, this suggests that information is lost through the categorisation of a continuous variable.

Table 5.15: Poisson regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = number of hospital visits (with BMI as discrete variables)

	(1)		(2)		(3)	
	β	SE	β	SE	β	SE
Hospital Visits (Equation 4.38)						
BMI Variables						
BMI1	-0.3607***	0.0854			-0.3998***	0.0859
BMI2	0.3768***	0.0890			0.3126***	0.0911
BMI3	-0.0292	0.3584			-0.1735	0.3624
Risk Variables						
DM_TY2			0.2791	0.2370	0.2377	0.2383
HYPER			0.1496	0.1044	0.1149	0.1055
DEPRESS			1.1859***	0.1210	1.1953***	0.1209
CHD			1.1296***	0.1219	1.1536***	0.1215
CHOLEST			-0.3351**	0.1393	-0.3378**	0.1389
MS_DIS			0.2991***	0.0770	0.2891***	0.0774
LL	-4538.81		-4480.18		-4455.19	
DF	45		48		51	
ML Ratio Test	--		--		(1) vs. (3)	$p < 0.01$
	--		--		(2) vs. (3)	$p < 0.01$

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood.

Looking at models 2 and 3, the following risk factors are statistically significant at the 5 per cent level or better: depression, CHD, elevated cholesterol levels, and musculoskeletal disorders. The results for model 3 are particularly interesting in that inclusion of the risk factor variables as a group is statistically significant. This is supported by the ML ratio test comparing models 2 and 3 ($p < 0.01$). Moreover, there is evidence to suggest that BMI exerts an independent impact on the expected number of hospital visits. However, the issue still remains that coefficients BMI1 and BMI3 in models 1 and 3 are negative and this finding is not consistent with the other types of MCU discussed earlier. Therefore, Equation (4.38) was re-estimated using BMI as continuous measure. These results are shown in Table 5.16 below.

Table 5.16: Poisson regression estimates (β) and standard errors (SE), (n = 28376); dependent variable = number of hospital visits (with BMI as a continuous variable)

	(1)		(2)		(3)	
	β	**SE**	β	**SE**	β	**SE**
Hospital Visits (Equation 4.38)						
BMI Variable						
BMI	0.0129*	0.0075			0.0062	0.0077
Risk Variables						
DM_TY2			0.2791	0.2370	0.2601	0.2382
HYPER			0.1496	0.1044	0.1370	0.1055
DEPRESS			1.1869***	0.1210	1.1849***	0.1210
CHD			1.1296***	0.1219	1.1328***	0.1219
CHOLEST			-0.3351**	0.1393	-0.3390**	0.1393
MS_DIS			0.2991***	0.0770	0.2936***	0.0773
LL	9126.14		8960.35		8959.72	
DF	43		48		49	
ML Ratio Test	--		--		(1) vs. (3)	$p < 0.01$
	--		--		(2) vs. (3)	NS

* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Notes: LL = log likelihood. DF = degrees of freedom. ML = maximum likelihood.

The results presented in Table 5.16 provide a better estimate of the relationship between the expected number of hospital visits and BMI. Model 1, in Table 5.16, shows that there is a positive statistical relationship between BMI and MCU, although the relationship is only statistically significant at the 10 per cent level. The corresponding BMI coefficient in model 1 can be interpreted as follows: a one-unit increase in BMI will lead to a 1.3 per cent increase in the average number of hospital visits.

Models 2 and 3 indicate that the following RF variables are statistically significant at the 5 per cent level of better: depression, CHD, elevated cholesterol levels, and musculoskeletal disorders. Moreover, when comparing models 1 and 3, it is evident from the corresponding ML ratio test that the introduction of the RF variables, as a

group, are statistically significant at the 1 per cent level. The BMI coefficient, in model 3, is relatively close to zero ($\beta = 0.0062$) and is not statistically significant as indicated by the ML ratio test comparing models 2 and 3. Once again, this indicates that the damaging impact of obesity is exerted through the associated risk factor variables.

5.6 Summary of Findings

This chapter has presented the results from the econometric models outlined in chapter 4. The major findings reported in this chapter are listed below:

(1) Overall, there is a statistically significant positive relationship between MCU and the level of obesity (as represented by '*c*' in Figure 4.1). Specifically, the following associations are indicated:

- Likelihood of a doctor visit increases with BMI;
- Likelihood of a visit to a health care professional increases with BMI;
- Likelihood of a hospital visit is *not* associated with BMI;
- *Expected number* of doctor visits increases with BMI;
- *Expected number* of visits to health care professional increases with BMI; and
- *Expected number* of hospital visits increases with BMI.

(2) Overall, the introduction of the six obesity-related risk factor variables significantly attenuated the independent impact of BMI variables, suggesting that increased MCU in these groups is related to these risk factors. In other words, obesity exerts its major impact through these six risk factors. For the most part, the ML ratio tests indicate that there is only little evidence to support the independent impact of BMI.

(3) The continuous measure of BMI generates a better fit. Overall the findings are consistent with the working hypotheses and current opinion in the medical literature.

These results presented in this chapter were also used to examine the policy implications and potential cost savings associated with a reduction in BMI. Specifically, those Poisson estimates that used BMI as a continuous measure will form the basis of the policy analysis in the next chapter.

CHAPTER 6: POLICY IMPLICATIONS

6.1 Introduction

The results presented in the previous chapter indicate that obesity exerts its damaging impact through a number of serious medical conditions. The purpose of this chapter is to utilise the results in chapter 5 to estimate the association between the direct cost of MCU and the level of BMI. These results were used to estimate the change in total costs associated with a one-unit change in BMI for people with a particular set of personal characteristics. Policy scenarios were also examined to estimate the potential cost savings associated with a reduction in body size (BMI) for various groups in the population.

6.2 A General Framework for Policy Analysis

The results presented in chapter 5 show that MCU and BMI are related. Information on this relationship was used to estimate the potential direct cost savings associated with a reduction in body mass. In examining these relationships, it is important to note that these estimates were based on the assumption of *reversibility* – that is, reductions in BMI could affect long-term, obesity-related risk factors (such as type 2 diabetes and hypertension) and, hence, the cost of MCU.

The evidence supporting reversibility is found in the medical literature. As discussed in chapter 2, there is a growing body of evidence indicating that a reduction in body weight is associated in a reduction in the mortality rate (e.g., Williamson et al. 1995). Moreover, there are a number of studies indicating that obesity-related risk factors such

as type 2 diabetes, hypertension, and elevated cholesterol levels are reduced by weight loss (refer to chapter 2, section 2.5.5). Therefore, there is a medical basis for assuming reversibility in the BMI, MCU relationship.

In terms of policy analysis, the total relationship (i.e., relationship *c* as depicted in figure 4.1 from chapter 4) was used to examine the association between the total direct cost of MCU and the level of BMI (obesity) using the Poisson estimates from chapter 5. The focus is on direct costs (i.e., the resource costs of providing medical care) rather than on the indirect costs, for example, lost productivity. The Poisson estimates presented in Table 6.1 below were used in the policy analysis presented below.

Table 6.1: Summary of Poisson regression estimates used in the policy analysis

Equation	*β*	*SE*	*Interpretation*
Doctor Visits (4.36)	0.0153***	0.0025	A one-unit increase in BMI will lead to a 1.5% increase in the mean number of doctor visits.
OHP Visits (4.37)	0.0212***	0.0033	A one-unit increase in BMI will lead to a 2% increase in the mean number of visits to other health care professionals
Hospital Visits (4.38)	0.0129*	0.0075	A one-unit increase in BMI will lead to a 1.3 % increase in the mean number of hospital visits.

Source: Tables 5.12, 5.14, and 5.16 from chapter 5.
* $p < 0.1$. ** $p < 0.05$. *** $p < 0.01$
Note: Continuous measure of BMI is used and estimates of control variables are contained in Appendix C on CD-ROM.

The β estimates presented in Table 6.1 were selected for the following reasons. First, the estimated coefficients are for a continuous measure of BMI. A continuous measure is preferable to a discrete measure, as information is not lost through categorisation. Second, Poisson regression estimates provide information on the *expected* or *mean* amount of medical care utilisation. This is important information as unit cost data can be attached to each type of medical care. The cost for each type of medical care can then be aggregated to provide an estimate of the total cost of MCU for this sample and

the population. Furthermore, information on total cost can then be used to assess the medical utilisation cost implications of a one-unit increase or decrease in BMI for the average person or a member of a given population segment. Information on the marginal cost of BMI in different segments of the population gives the policy maker a guide as to where to focus health care resources to reduce the medical costs associated with obesity.

6.3 Sources of Cost Data

For this project, three types of medical care utilisation measures are used: doctor visits; other health care professional visits (OHP); and hospital visits. To quantify the relationship between the cost of MCU and BMI, cost data were selected following a search of the literature. The objective of this search was to identify suitable cost data for each type of MCU. The cost data and their sources are listed in Table 6.2 below.

Table 6.2: Cost data and their sources

Item	*Cost*	*Source*
Standard General Practitioner consultation	$26.45	Medicare Benefits Schedule, 2000 (1 November 1999 Including 1 February 2000 and 1 May 2000 Supplements)
Average cost of OHP consultation	$48.15	Schedule for Allied Health Fees, 1999
Average Hospital Visit Cost, Australia	$2,529	Australia's Health, 1998

Notes: Complete details on the current Medicare Benefits Schedule (MBS) can be found at the following website: http://www.health.gov.au/pubs/mbs/mbs6/default.htm. This page was produced by Health Access and Financing Division, Commonwealth Department of Health and Aged Care, 14 March 2000.

The standard general practitioner (GP) schedule fee of $26.45 was selected from the *Medicare Benefits Schedule* (2000) to represent the cost associated with a doctor visit.[6]

[6] Note that it will be an under-estimate insofar as many GPs charge a higher fee. Against this, there is much bulk billing, where the effective fee charged is only 85 per cent of the schedule fee. Accordingly, the schedule fee is considered a reasonable estimate of the direct costs of GP visits.

This cost assumption was selected after seeking advice from the Health Access and Financing Division at the Commonwealth Department of Health and Aged Care. A similar cost assumption was also used in a study by Dalton et al. (1997) which examined the cost-effectiveness of GP-led behavioural change involving weight reduction.

The cost of an OHP consultation was estimated using a simple average of the costs associated with the following types of visits to a dietician ($30.40 per half-hour), physiotherapist ($39.40), podiatrist ($33.90), psychologist ($85.90), speech pathologist ($59.70), and social worker ($39.65). These costs were obtained from the *Consultation and Treatment Fees for Allied Health Practitioners* (1999) from the Health Access and Financing Division, Commonwealth Department of Health and Aged Care.

The average hospital cost estimate of $2,529 per instance of use was selected from the Australian Institute of Health and Welfare's report entitled *Australia's Health* (1998). This cost is an indicator that ". . . measures the average cost of providing care for an admitted patient (whether an overnight-stay patient or a same-day patient) adjusted for the relative complexity of the patient condition and of the hospital services performed" (Australia Institute of Health and Welfare, 1998, p. 191). The cost components contained in this measure take into account medical labour, non-medical labour, and indirect costs.

6.4 Marginal Cost of MCU as a Function of BMI

The Poisson regression estimates for the three different types of MCU (Table 6.1) and the unit cost assumptions (Table 6.2) were used to quantify the cost of increased MCU associated with increased BMI. (This is henceforth referred to as 'the marginal cost of BMI' for convenience.) The cost associated with each type of MCU was aggregated to provide an overall measure of the marginal cost of BMI. For aggregation purposes, it was assumed that there was independence between the three different types of MCU (e.g., a visit to a doctor was independent of a visit to another health care professional and so on). This assumption was based on the structure of the survey, in which respondents were specifically asked about their MCU in the 2 weeks prior to the survey. Owing to the survey design, it was reasonable to assume that there was independence between the number of doctor visits, other health care professional visits, and hospital visits. Taking into account the cost data and Poisson regression estimates, the marginal cost of MCU was estimated for individuals with a particular set of personal characteristics at a given level of BMI.

6.4.1 A Framework for Generating Marginal Cost Estimates

As previously noted, Table 6.1 provides estimates of the impact of BMI on the expected number of visits to doctors, other health care professionals, and hospitals. Table 6.2 provides a list of the cost assumptions and their sources. Together, this information can be used to estimate the marginal cost of BMI for an individual with a particular set of personal characteristics at a given level of BMI.

To begin with, consider Equation (6.1) below which provides a definition of the expected total cost of MCU for an individual:

$$TC = C_D\hat{N}_D + C_O\hat{N}_O + C_H\hat{N}_H \tag{6.1}$$

where:

TC = expected total cost of medical care utilisation (MCU);

$\hat{N}_D$ = expected number of doctor visits;

$\hat{N}_O$ = expected number of OHP visits;

$\hat{N}_H$ = expected number of hospital visits;

C_D = cost of doctor visit ($26.45);

C_O = cost of other health care professional visit ($48.15); and

C_H = cost of hospital day ($2,529).

The marginal cost (MC) associated with a one-unit change BMI is shown below:

$$\frac{\partial TC}{\partial(BMI)} = C_D\frac{\partial\hat{N}_D}{\partial(BMI)} + C_H\frac{\partial\hat{N}_H}{\partial(BMI)} + C_O\frac{\partial\hat{N}_O}{\partial(BMI)} \tag{6.2}$$

Now, consider the estimated number of doctor visits, obtained from the Poisson regression (Equation 4.36 from Table 6.1):

$$\hat{N}_D = exp(\hat{\theta}_0)\cdot exp(\hat{\theta}_1\cdot BMI) \tag{6.3}$$

where:

$exp(\hat{\theta}_0)$ captures the effect of all control variables such as age, sex, and so on.

Differentiating (6.3) with respect to BMI gives:

$$\frac{\partial \hat{N}_D}{\partial (BMI)} = exp(\hat{\theta}_0) \cdot \hat{\theta}_1 exp(\hat{\theta}_1 \cdot BMI) = \hat{\theta}_1 \hat{N}_D \quad (6.4)$$

That is, the MC of BMI associated with the expected number of doctor visits is:

$$MC_D = \hat{\theta}_1 C_D \hat{N}_D \quad (6.5)$$

This procedure can also be applied to the expected number of other health care professional visits and hospital visits. Thus, the marginal cost of BMI is defined as:

$$MC = \hat{\theta}_1 C_D \hat{N}_D + \hat{\theta}_2 C_O \hat{N}_O + \hat{\theta}_3 C_H \hat{N}_H \quad (6.6)$$

Equation (6.6) provides a framework to generate marginal cost of BMI estimates for individuals. Given the vector of socioeconomic characteristics (i.e., control variables) and BMI, the expected number of medical care visits was determined. Given the coefficient estimates, the effect on cost of MCU of a unit increase in BMI for this type of person can be estimated from (6.6). With reversibility, these MC estimates also define the marginal benefits of weight loss in terms of direct cost savings. As previously indicated, the Poisson regression estimates presented in Table 6.1 and the cost data from Table 6.2 were used to estimate the marginal costs for BMI. The following section provides marginal cost examples for individuals with a particular set of personal characteristics at selected BMI values.

6.4.2 Marginal Cost Estimates: Selected Examples

Using this framework, it is now possible to quantify and examine the relationship between the marginal cost of BMI at different levels of BMI. For expository purposes, the estimated Poisson regression coefficients (as reported in Table 6.1) were selected

(refer to Tables C2, C7, and C12 in Appendix C for complete details of all estimated coefficients *including control variables*). Of the eight different types of control variables, a number of different categories were statistically significant. However, age and sex were the most statistically significant control groups. For purely illustrative purposes however, all the information contained within Tables C2, C7, and C12 was used to provide an estimate of the marginal cost of BMI conditional upon a set of personal characteristics at a given level of BMI. Four examples are provided below to illustrate how to derive marginal costs estimates.

Person Type 1

For person type 1, the following characteristics were selected: male, age 45-49, with no higher qualifications, born in Australasia, earning a wage between $20,000-$24,999, residing in a capital city with no private health insurance cover (the appropriate control variable coefficients were selected from Tables C2, C7, and C12). For an individual with these personal characteristics and a BMI value of 20, the expected marginal cost of MCU (over the 2-week survey period) is:

MC_1 = [\$26.45*0.0153**exp*(-1.4777 + 0.0153*20)] + [\$48.15*0.0212**exp*(-2.4426 + 0.0212*20)] + [\$2529*0.0129**exp*(-5.0444 + 0.0129*20)

= \$0.53

In other words, a one-unit reduction in BMI will result in a \$0.53 reduction in MCU costs in a 2-week period. This figure can be multiplied by 26 weeks to estimate the annual MC associated with a one-unit change in BMI (i.e., \$0.53*26 = \$13.86).

Holding these personal characteristics fixed, the marginal cost of MCU can also be estimated at various levels of BMI. Table 6.3 below shows the value of the marginal cost of MCU at different levels of BMI.

Table 6.3: Marginal cost of BMI for person type 1

BMI	*MC ($) – 2 week period*	*MC ($) – per year*
18.5	0.52	13.54
20	0.53	13.86
25	0.58	14.99
30	0.63	16.34
35	0.67	17.53
40	0.73	18.97

Note: Marginal Cost (MC) estimates are for men, age 45-49, with no higher qualifications, born in Australasia, earning a wage between $20,000-$24,999, residing in a capital city with no private health insurance cover. Differences between fortnightly and yearly estimates are due to rounding.

This example highlights that marginal cost rises (at an increasing rate) as BMI increases [as can be confirmed from differentiation of (6.6)]. This is to be expected as higher levels of BMI are associated with greater resource utilisation in the medical care sector. Therefore, a reduction in BMI will result in cost savings. For example, a man with the above characteristics and a BMI value of 40 will save $18.97 per year of medical costs as a result of a one-unit reduction in BMI. To explore these issues more fully, extensions of this example are presented below.

Person Type 2

Person type 2 is defined as female age 45-49 with no higher qualifications, born in Australasia, earning a wage between $20,000-$24,999, residing in a capital city and with no private health insurance cover. The only difference in this example is that the individual is female. Therefore, taking this into account, the expected marginal cost of MCU for a female with the above set of characteristics and a BMI value of 20 is:

MC_2 = [\$26.45*0.0153*exp(-1.7917 + 0.0153*20)] + [\$48.15*0.0212*exp(-2.7471 + 0.0212*20)] + [\$2529*0.0129*exp(-5.2124 + 0.0129*20)

= \$0.42

Table 6.4 below shows the relevant MC estimates for selected BMI values.

Table 6.4: Marginal cost of BMI for person type 2

BMI	*MC (\$) – 2 week period*	*MC (\$) – per year*
18.5	0.41	10.71
20	0.42	10.96
25	0.46	11.84
30	0.49	12.80
35	0.53	13.83
40	0.57	14.95

Note: Marginal Cost (MC) estimates are for women, age 45-49, with no higher qualifications, born in Australasia, earning a wage between \$20,000-\$24,999, residing in a capital city with no private health insurance cover. Differences between fortnightly and yearly estimates are due to rounding.

Interestingly, men with identical characteristics (person type 1, Table 6.3) have a higher MC profile when compared to women (person type 2, Table 6.4). These examples suggest that greater potential cost savings could be obtained by targeting overweight and obese men, other things equal.

Person Type 3

Person type 3 is defined as male, age 55-59, with no higher qualifications, born in Australasia, earning a wage between \$20,000-\$24,999, residing in a capital city with no private health insurance cover. Except for the selected age group (55-59), the following set of characteristics is same as for person type 1. The MC for a man with these characteristics and a BMI value of 20 is:

MC_3 = [\$26.45*0.0153*exp(-1.3759 + 0.0153*20)] + [\$48.15*0.0212*exp(-2.2592 + 0.0212*20)] + [\$2529*0.0129*exp(-5.1013+ 0.0129*20)

= \$0.52

The MCU estimates for different BMI values are presented in Table 6.5 below.

Table 6.5: Marginal cost of BMI for person type 3

BMI	*MC (\$) – 2 week period*	*MC (\$) – per year*
18.5	0.50	13.09
20	0.52	13.40
25	0.56	14.48
30	0.60	15.65
35	0.65	16.92
40	0.70	18.30

Note: Marginal Cost (MC) estimates are for men, age 55-59, with no higher qualifications, born in Australasia, earning a wage between \$20,000-\$24,999, residing in a capital city with no private health insurance cover. Differences between fortnightly and yearly estimates are due to rounding.

Not surprisingly, these MC estimates are similar to the estimates presented in Table 6.3 for person type 1 though slightly lower over the range of BMI.

Person Type 4

Person type 4 is defined as female, age 55-59, with no higher qualifications, born in Australasia, earning a wage between \$20,000-\$24,999, residing in a capital city with no private health insurance cover. These characteristics are the same as for person type 2, but older and differ from those of person type 3 in being female.

MC_4 = [\$26.45*0.0153*exp(-1.6899 + 0.0153*20)] + [\$48.15*0.0212*exp(-2.8733 + 0.0212*20)] + [\$2529*0.0129*exp(-5.2693 + 0.0129*20)

= $0.41

Table 6.6 below lists the MC of MCU for different values of BMI for person type 4. These results are consistent with those given previously. The marginal costs for women in this group are uniformly lower than those for equivalent males and lower than those for those for younger women.

Table 6.6: Marginal cost of BMI for person type 4

BMI	*MC ($) – 2 week period*	*MC ($) – per year*
18.5	0.40	10.34
20	0.41	10.58
25	0.44	11.42
30	0.47	12.33
35	0.51	13.32
40	0.55	14.40

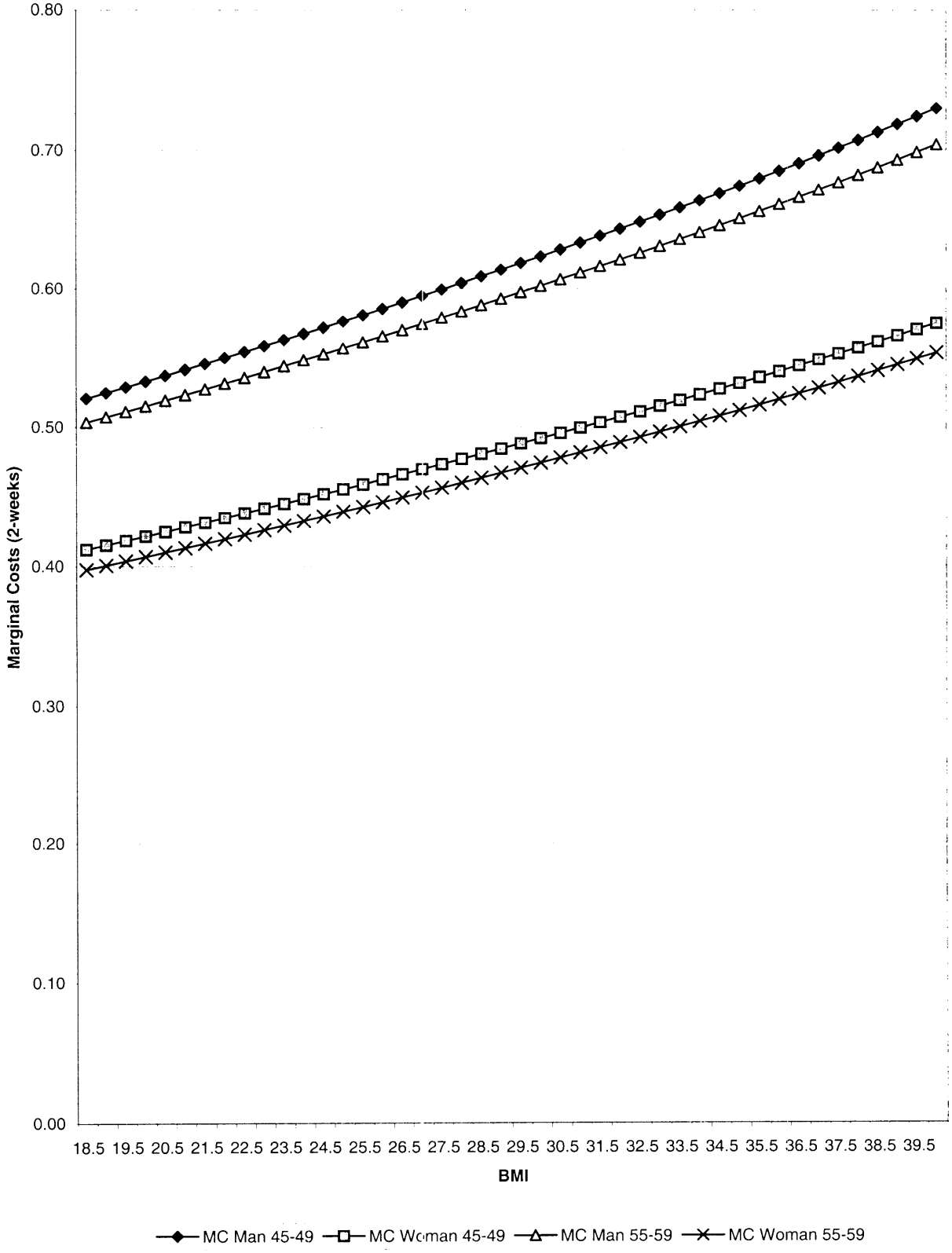
Notes: Marginal Cost (MC) estimates are for women, age 55-59, with no higher qualifications, born in Australasia, earning a wage between $20,000-$24,999, residing in a capital city with no private health insurance cover. Differences between fortnightly and yearly estimates are due to rounding.

6.4.3 Discussion of Selected Examples

The examples presented above demonstrate how marginal cost estimates can be calculated for individuals with a certain set of personal characteristics at different levels of BMI. In order to facilitate the discussion, the above MC estimates (for the 2-week period) were plotted against BMI in Figure 6.1 below. Figure 6.1, shows the relationship between the MC of MCU and the level of BMI for person type 1 to person type 4. Taking into account these relationships, it is worth noting that:

- There is a positive relationship between MC and the level of BMI; and
- The difference between the MC profiles is relatively small.

Figure 6.1: Marginal cost of MCU



It is possible to generate a whole family of MC cost profiles dependent on the particular selection of the person characteristics (i.e., the estimated control variable coefficients presented in the appropriate tables in Appendix C). In fact, there are *numerous* combinations of personal characteristics that could be selected.

The difference between the MC profiles shown in Figure 6.1 is relatively small, which indicates that the fine level of detail is not really necessary. Although these examples provide a framework for estimating MC, the real challenge is to provide MC estimates for various segments of the Australian *population* aged 20 to 64.

In order to provide an estimate of the MC for segments of the Australian population, it is necessary to reduce the amount of detail contained within the extensive number of control variable categories. Ideally, it would be preferable to aggregate all MC estimates across *all* the different types of people as defined by the control variables. However, this approach is not practical given the level of detail presented in the above examples. Therefore, to generate MC estimates for various segments of the population, it is necessary to drastically reduce the number of control variable categories.

6.5 Marginal Cost Estimates for Segments of the Population

To obtain MC estimates for segments of the Australian population (aged 20-64), it was necessary to select personal characteristics that were manageable. After considering *all* the relevant Poisson coefficients (as summarised in Table 6.1) a decision was made to select only two of the control variables namely; age and sex. This decision was made for the following reasons: (i) overall, the age and sex coefficients were statistically

significant, (ii) the other categories do not contribute very much to describing the data, and (iii) it was a convenient approach in terms of broadly summarising (and categorising) the Australian population.

Taking this into account, Equations (4.36), (4.37), and (4.38) were then re-estimated using the following independent variables: age (re-coded into three broad categories), sex, and a continuous measure of BMI. The age variable was grouped into the following categories: (i) 20-39, (ii) 40-54, and (iii) 55-64 and two discrete variables were used, with category (i) selected as the excluded reference group. Summarising the sample according to the sex and age variables results in the following six categories:

- Women 20-39;
- Women 40-54;
- Women 55-65;
- Men 20-39;
- Men 40-54; and
- Men 55-64.

The above six categories provide a simplified framework necessary to generate MC estimates for these particular segments of the Australian population. The re-estimated Poisson regression equations are shown in Table 6.7 below. The BMI coefficients, as shown in Table 6.7, are highly significant for all equations ($p < 0.01$). Furthermore, all but one of the control variables are statistically significant at the 1 per cent level. These estimated coefficients were then used to provide an estimate of the MC of MCU for each of the six categories listed above.

Table 6.7: Poisson regression estimates (β) and standard errors (SE), (n = 28376); dependent variables = MCU (re-estimated)

	β	*SE*	*Label*
Doctor Visits (Equation 4.36)			
Control Variables			
AGE1	-0.0273	0.0247	40-54
AGE2	0.1387***	0.0152	55-64
SEX	-0.4352***	0.0222	Female
BMI Variable			
BMI	0.0203***	0.0025	Continuous
Constant	-1.5621***	0.0629	
OHP Visits (Equation 4.37)			
Control Variables			
AGE1	-0.1180***	0.0321	40-54
AGE2	-0.0933***	0.0229	55-64
SEX	-0.3453***	0.0293	Female
BMI Variable			
BMI	0.0244***	0.0032	Continuous
Constant	-2.1731***	0.0832	
Hospital Visits (Equation 4.38)			
Control Variables			
AGE1	-0.2593***	0.0802	40-54
AGE2	0.2086***	0.0436	55-64
SEX	-0.5253***	0.0692	Female
BMI Variable			
BMI	0.0263***	0.0074	Continuous
Constant	-3.8932***	0.1911	

*** $p < 0.01$

The estimated coefficients for the control and BMI variables in Table 6.7 were combined with the unit cost data presented in Table 6.2 to generate marginal cost estimates for each segment of the population. Details of the MC calculations for women aged 20-39 are provided in the following section.

MC Estimates for Women Aged 20-39

In this section, a detailed account is presented of how to generate MC estimates for the population group: women aged 20-39. As the focus is on MC estimates for the Australian population it would be helpful to provide an overview of population estimates derived from the 1995 National Health Survey (NHS). In fact, the 1995 NHS contains weights (or expansion factors) for each record contained in the survey. These weights can then be used to ". . . provide estimates relating to the whole population within the scope of the survey" (Australian Bureau of Statistics, 1996, p.20).

From the sample used in this project (n = 28,376), it is possible to obtain relevant population estimates. Information on the profile of the population by sex, age group, and BMI are presented in Table 6.8 below. Several aspects of Table 6.8 are worth noting. Based on a sample of 28,376 records, it was estimated that 9,531,510 Australians were between the age of 20-64. Of this total, 52 per cent were female and 48 per cent were male. Interestingly, 4,738,842 individuals (or about 50 per cent of people aged 20-64) were classified as being either overweight or obese (i.e., a BMI > 25). This finding is consistent with recent US estimates, which indicate that one of every two adults is either overweight or obese (Must et al. 1999). Moreover, 1,303,152 (14 per cent) of people between 20-64 were classified as obese (BMI $\geq$ 30). This information is important as it can be used in conjunction with the Poisson regression estimates to calculate the MC for segments of the Australian population.

Table 6.8: Population estimates derived from the 1995 NHS by age, sex, and BMI

Population	*Number*
Sex	
Females 20-64	4,991,155
Males 20-64	4,540,355
Total	9,531,510
Population by Sex and Age Groups	
Female	
Age 20-39	2,596,049
Age 40-54	1,684,932
Age 55-64	710,174
Male	
Age 20-39	2,347,490
Age 40-54	1,544,789
Age 55-64	648,076
Population with a BMI ≥ 25 by Sex and Age Groups	
Female	
Age 20-39	1,324,540
Age 40-54	1,096,690
Age 55-64	476,842
Total Female	2,898,072
Male	
Age 20-39	738,415
Age 40-54	739,318
Age 55-64	363,037
Total Male	1,840,770
All Persons	4,738,842
Population with BMI ≥ 30 by Sex and Age Groups	
Female	
Age 20-39	287,128
Age 40-54	281,137
Age 55-64	123,635
Total Female	691,900
Male	
Age 20-39	240,071
Age 40-54	250,085
Age 55-64	121,096
Total Male	611,252
All Persons	1,303,152

Note: These population figures are based on the population weights from the sample of 28,376 records selected from the 1995 National Health Survey.

Table 6.9 below, provides details on the selected Poisson regression coefficients for women aged 20-39. For convenience, the β coefficients are reproduced in the second column of Table 6.9.

Table 6.9: Poisson coefficients for women aged 20-39

Doctor Visits Equation (4.36)	β
AGE1	-0.0273
AGE2	0.1387
SEX	-0.4352
BMI	0.0203
Constant	-1.5621
OHP Visits Equation (4.37)	
AGE1	-0.1180
AGE2	-0.0933
SEX	-0.3453
BMI	0.0244
Constant	-2.1731
Hospital Visits Equation (4.38)	
AGE1	-0.2593
AGE2	0.2086
SEX	-0.5253
BMI	0.0263
Constant	-3.8932

To estimate the MCU for females in the 20-39 age group (i.e., the reference age group) at a particular BMI, AGE1 and AGE2 are set equal to zero and the remaining coefficient estimates are substituted into the MC function (6.6), along with the relevant cost estimate and specified level of BMI. For example, the (two-week) marginal cost for women in this group with a BMI of 20 is given by:

$$MC_{Female20\text{-}39} = [\$26.45*0.0203*exp(-1.9972 + 0.0203*20)] + [\$48.15*0.0244*exp(-2.5184 + 0.0244*20)] + [\$2529*0.0263*exp(-4.4186 + 0.0263*20)]$$

= $1.62

The procedure for calculating the above MC is same method employed in section 6.4.2. As before, the MC can be calculated at different levels of BMI. Table 6.10 below provides complete details for the MC profile of women in the population aged 20-39. Column (1) lists the MC per person per year. Column (2) lists the body mass index (BMI) groups. For convenience, these BMI groups were calculated at the midpoint. For example, all individuals whose BMI values were between 24.75 and 25.25 were assigned a BMI value of 25. Column (3) provides information on the *number* of females in the *population* aged 20-39, who fall into each BMI group. For example, a total of 49,163 women are classified in the BMI group of 30. The sum of column (3) is 2,596,049, which is the total number of females in the population aged 20-39 (as reported in Table 6.8). Finally, column (4) provides an estimate of the potential cost savings per year. These estimates were calculated by multiplying column (1) by column (3). That is, the MC per person per year by the *number* of women in each BMI category.

For illustrative purposes, consider the following example (from Table 6.10 above) for women aged 20-39 with a BMI value of 30. The potential cost savings associated with this particular group of women is equal to $2,677,510 per year. This can be interpreted as follows: if *all* women in this group reduced their BMI value by one-unit, a potential annual cost saving of about $2.7 million would be realised for the year.

Table 6.10: MC profile and potential cost savings for women aged 20-39

(1)	*(2)*	*(3)*	*(4)*
Per year	*BMI*	*Number in population*	*Potential Cost Saving*
$47.89	25	177,608	$8,505,314
$48.51	25.5	146,612	$7,111,870
$49.14	26	132,394	$6,505,306
$49.77	26.5	99,813	$4,967,881
$50.42	27	114,468	$5,771,076
$51.07	27.5	78,279	$3,997,658
$51.73	28	78,642	$4,068,182
$52.40	28.5	100,750	$5,279,296
$53.08	29	43,580	$2,313,189
$53.77	29.5	65,266	$3,509,081
$54.46	30	49,163	$2,677,510
$55.17	30.5	29,587	$1,632,261
$55.88	31	36,873	$2,060,523
$56.61	31.5	30,769	$1,741,725
$57.34	32	21,070	$1,208,160
$58.08	32.5	20,700	$1,202,311
$58.83	33	19,986	$1,175,834
$59.60	33.5	13,544	$807,186
$60.37	34	11,175	$674,614
$61.15	34.5	9,895	$605,052
$61.94	35	6,409	$397,017
$62.75	35.5	8,608	$540,134
$63.56	36	5,693	$361,854
$64.38	36.5	4,509	$290,275
$65.22	37	3,811	$248,531
$66.06	37.5	2,893	$191,133
$66.92	38	2,109	$141,119
$67.78	38.5	870	$58,954
$68.66	39	977	$67,104
$69.55	39.5	1,068	$74,276
$70.45	40	1,184	$83,441
$71.37	40.5	1,838	$131,169
$72.29	41	771	$55,738
$73.23	41.5	2,157	$157,973
$74.18	42	435	$32,249
$75.14	42.5	47	$3,563
$76.11	43	0	$0
$77.10	43.5	985	$75,970

The information presented in Table 6.10 can also be used to examine the overall cost savings for women in the population who are (i) either overweight/obese (i.e., BMI $\geq$ 25) or (ii) obese (BMI $\geq$ 30). For instance, the potential cost savings for all women who

are either overweight or obese is about $68.7 million per annum. For obese women, the corresponding amount is $16.7 million per annum. Thus, if all women in the population with a BMI value ≥ 25 reduced their BMI value by one-unit a potential cost saving of $68.7 million per annum would be realised. For obese women, as a group, a one-unit reduction would result in a cost saving of about $16.7 million. These amounts indicate that the cost savings associated with a reduction in body size are substantial.

Potential Cost Savings for the Remaining Segments of the Population

The above section provided a detailed account of how to generate MC estimates for women aged 20-39. In addition, potential cost savings were also presented for sections of this particular segment of the population. The procedure presented above was also used to generate MC profiles and potential cost savings for all six specified categories of the Australian population. Table 6.11 below provides a summary of the relationship between MC and BMI for these segments of the population. It should be noted that the selected BMI values are presented for comparative purposes.

Table 6.11 shows yearly MC estimates for the six segments of the Australian population. A range of BMI values was chosen to highlight the different MC profiles for women and men in the three different age groups. As described in the previous section above, the yearly MC estimates for all six categories of the Australian population were multiplied by the population distribution to provide an estimate of the potential cost savings. In other words, the same procedure as presented in Table 6.10 can be used to estimate the potential cost savings for each segment of the population.

Table 6.11: Marginal cost per person per year by BMI

Age and Sex by BMI value	*MC per person per year*
BMI 25	
Women 20-39	$47.89
Women 40-54	$38.11
Women 55-64	$57.28
Men 20-39	$79.26
Men 40-54	$62.89
Men 55-64	$95.18
BMI 30	
Women 20-39	$54.46
Women 40-54	$43.32
Women 55-64	$65.16
Men 20-39	$90.16
Men 40-54	$71.50
Men 55-64	$108.31
BMI 35	
Women 20-39	$61.94
Women 40-54	$49.25
Women 55-64	$74.13
Men 20-39	$102.57
Men 40-54	$81.30
Men 55-64	$123.25
BMI 40	
Women 20-39	$70.45
Women 40-54	$55.99
Women 55-64	$84.34
Men 20-39	$116.70
Men 40-54	$92.45
Men 55-64	$140.25

Details of the potential cost savings are shown in Table 6.12 below. The potential cost savings, presented in Table 6.12, indicate that a reduction in body size could lead to substantial cost savings related to MCU. The magnitude of potential savings is greatest for overweight/obese men and women in the 20-39 year age group. For instance, the expected cost savings associated with a one-unit reduction in BMI for all men aged 20-39 with a BMI $\geq$ 25 is equal to about $65 million per year. For obese men, in this age group, the expected cost savings associated with a one-unit reduction in BMI are

estimated to be in the order of $25 million. For all persons aged 20-64 in the population, the expected cost savings are substantial.

Table 6.12: Potential cost savings for segments of the population

	Women		*Men*		*All Persons*	
Age	***BMI ≥ 25***	***BMI ≥ 30***	***BMI ≥ 25***	***BMI ≥ 30***	***BMI ≥ 25***	***BMI ≥ 30***
20-39	$68,724,530	$16,695,676	$65,417,243	$24,047,465	$134,141,773	$40,743,142
40-54	$45,573,473	$12,907,376	$52,083,788	$19,790,674	$97,657,261	$32,698,049
55-64	$29,855,447	$8,577,059	$38,667,017	$14,456,384	$68,522,465	$23,033,443
Total	*$144,153,450*	*$38,180,111*	*$156,168,048*	*$58,294,523*	*$300,321,499*	*$96,474,634*

For instance, if all people aged 20-64 with a BMI ≥ 25 reduced their BMI by one-unit, a potential cost saving of about $300 million per annum could be realised. These cost estimates indicate that significant savings could be achieved in the medical sector through a reduction in body size. However, while these results provide estimates of the MC for segments of the Australian population, it is important to demonstrate the magnitude of potential cost savings through simple policy scenarios. The following section discusses the results from three policy scenarios. These scenarios were selected to demonstrate the potential cost savings that could be realised through weight reduction and maintenance strategies.

6.6 Policy Scenarios

Three policy scenarios were conducted to assess the potential cost savings associated with a reduction in BMI: (i) shifting all overweight/obese individuals to a 'healthy' BMI value of 25; (ii) reducing obese individuals' BMI value by 5 per cent (i.e., all

individuals' with a BMI ≥ 30); and (iii) reducing the BMI of all obese individuals by 10 per cent.

The first policy scenario is referred to as an *ideal* outcome. The purpose of this scenario is to estimate the potential cost savings with shifting all overweight/obese individuals to a BMI value of 25. It provides an indication of the magnitude of the potential cost savings if overweight and obesity could be eliminated in all persons aged 20-64 in the Australian population. Equivalently, it measures the total direct MCU cost of obesity and overweight.

The total cost equation (6.1) from section 6.4.1 above was used to estimate the potential cost savings associated with this ideal policy scenario. For illustrative purposes, consider the following example, which highlights the potential cost savings associated with reducing all women aged 20-39 from a BMI value of 30 to a BMI value of 25. Table 6.13 below provides details on how this policy scenario was calculated.

In Table 6.13, the following information is presented. The selected BMI values for women aged 20-39, the total number of individuals in this segment of the population, the total cost (TC) per person, and the TC per person per year. Thus the estimated annual savings in MCU costs if a woman aged 20-39 with a BMI of 30 were to reduce her weight to attain a 'healthy' BMI of 25 would be $2,125 - $1,869 = $256 per annum.

Table 6.13: Shifting women aged 20-39 from a BMI of 30 to 25

BMI	*No. in population*	*TC per person*	*TC per year/per person*
25	177,608	$71.92	$1,869.90
30	49,163	$81.75	$2,125.43

As previously indicated, the TC estimates were calculated using Equation (6.1). For example, the TC per person for women with a BMI equal to 30 was calculated as follows:

$$TC_{Female20\text{-}39} = [\$26.45*exp(-1.9972+ 0.0203*30)] + [\$48.15*exp(-2.5184 + 0.0244*30)] + [\$2529*exp(-4.4186 + 0.0263*30)]$$

$$= \$81.75$$

This figure corresponds to an annual estimate of $2,125.43. Therefore, the difference between the yearly TC figures provide an estimate of the potential cost savings per person per year. In this example, the saving per person per year is equal to $255.52. Multiplying this amount by the total number of women results in an aggregate saving of about $12.6 million. The approach described above was applied to each segment of the Australian population to generate an estimate of potential savings that could be realised if overweight and obesity could be eliminated in all persons aged 20-64. The results from this ideal policy scenario are presented in Table 6.14 below.

The estimates presented in Table 6.14 indicate that under this ideal policy scenario, the potential cost savings are substantial. For all persons, this scenario suggests that if

overweight and obesity were eliminated, savings of about $1 billion could be achieved. To put this policy scenario into perspective, this saving represents approximately 2 per cent of total health care expenditure in 1997-98. During this period it was reported that total health expenditure totalled about $47 billion or 8 per cent of gross domestic product (Australian Institute of Health & Welfare, 2000).

Table 6.14: Policy scenario 1 – Shifting individuals to a BMI value of 25

Age and Sex	*Potential Annual Cost Savings*
Women	
20-39	$205,786,406
40-54	$147,268,248
55-64	$98,593,799
Subtotal	$451,648,453
Men	
20-39	$267,253,553
40-54	$217,701,671
55-64	$159,150,976
Subtotal	$644,106,199
Total All Persons	$1,095,754,652

From a health policy perspective, the breakdown of the potential savings provides an indication of where the greatest benefits could be achieved. In this ideal scenario, 59 per cent of cost savings were attributed to men compared to 41 per cent for women. This indicates that a greater proportion of cost savings (in terms of medical care utilisation) could be achieved by specifically targeting overweight and obese men. It is also worth noting that for both men and women, the greatest savings are realised for the age group 20-39. For men 20-39, the potential cost savings is about $267 million per annum. The corresponding saving for women is about $206 million per annum. Noting that these are annual cost savings, any policy that successfully and permanently reduces

the weight of the young will have more long-term benefits than successful policies targeted at older population segments.[7]

The second policy scenario was designed to estimate the potential cost savings that could be realised by reducing BMI values (in those individuals with a BMI value ≥ 30). In other words, what are the cost saving associated with a 5 per cent reduction in the body mass of obese individuals? The purpose of this policy scenario was to estimate the benefits associated with a moderate (and achievable) reduction in body size. The estimates for the second policy scenario are displayed in Table 6.15 below.

Table 6.15: Policy scenario 2 – Reducing obese individuals' BMI values by 5%

Age and Sex	*Potential Annual Cost Savings*
Women	
20-39	$26,671,435
40-54	$20,437,659
55-64	$13,652,114
Subtotal	$60,761,208
Men	
20-39	$40,318,412
40-54	$33,063,282
55-64	$23,984,534
Subtotal	$97,366,229
Total All Persons	$158,127,437

The total cost approach as described above was also used to estimate the second policy scenario. As shown in Table 6.15, the potential cost saving associated with a 5 per cent

[7] A complete cost-effectiveness analysis would generate estimates of the present value of cost savings for different segments. This would clearly favour policies directed at the young. However, since the obese die earlier than the non-obese, these higher benefits may be partly or totally offset by the increased medical costs associated with a longer life span. These issues are not investigated here.

reduction in BMI for obese individuals was about $158 million per annum. Thus, a very modest reduction in BMI for all obese individuals aged 20-64 could result in substantial savings. Once again, greater cost savings can be achieved by targeting obese men. The potential cost savings for men was estimated to be about $97 million per annum. The corresponding cost savings for women was about $61 million per annum.

The third policy scenario was designed to estimate the potential cost savings from a 10 per cent reduction in BMI for obese individuals aged 20-64. The results for this policy scenario are presented in Table 6.16 below.

Table 6.16: Policy scenario 3 – Reducing obese individuals' BMI values by 10%

Age and Sex	*Potential Annual Cost Savings*
Women	
20-39	$52,239,427
40-54	$40,041,664
55-64	$26,740,383
Subtotal	$119,021,473
Men	
20-39	$78,869,139
40-54	$64,686,894
55-64	$46,929,988
Subtotal	$190,486,021
Total All Persons	$309,507,494

A 10 per cent reduction in BMI is reasonable – especially for obese individuals – in terms of both potential cost savings and the positive health improvement associated with weight loss. As shown in Table 6.16, if all persons aged 20-64 with a BMI $\geq$ 30 were able to reduce their BMI value by 10 per cent, potential cost savings of about $310 million per annum could be achieved. Consistent with the previous two policy

scenarios, the greatest economic benefits could be achieved by targeting obese men, especially those men aged 20-39. The information from these policy scenarios indicates that the potential cost savings associated with reducing overweight and obese individuals is substantial. The following section discusses these policy scenarios.

6.7 Discussion of Policy Scenarios

A good case can be made that these policy scenario figures underestimate the *true* potential cost savings associated with a reduction in the level of obesity for the following reasons. As indicated above, the cost of MCU comprises the sum of the costs of (i) the expected number of doctor visits, (ii) the expected number of visits to other health care professionals, and (iii) the expected number of hospital visits. The mean cost of MCU for the 1995 NHS sample is $868 per annum (i.e., the mean value of the cost data presented in Table 6.2 above). However, it should be noted that in 1997-98 per capita Australian health care expenditure was estimated at $2,523 (Australian Institute of Health & Welfare, 2000). These differences suggest that the total cost estimates are, in fact, biased downwards because only doctor visits, other health care professional visits, and hospital visits are included in the definition of MCU. The definition of the cost of MCU does not account for other medical costs associated such as other home health care, nursing home care, medical durables, drugs and other non-durables, and other personal care. Nevertheless, taking into account the limitations associated with the availability of suitable cost data, the above policy scenarios still provide a guide as to the overall size of the potential cost savings that could be achieved by implementing weight reduction strategies.

Secondly, these policy scenarios do not take into account the benefits associated with a reduction in indirect costs (e.g., a reduction in lost productivity). A case can be made that losses in productivity due to obesity-related illness, disability, and sudden death (Gorstein & Grosse, 1994) result in lost income by the patient and their family (Langley, 1993). Therefore, a reduction in indirect costs would be achieved through the implementation of weight loss strategies, especially among obese members of society.

The policy scenarios are presented in terms of a reduction in body size (BMI) as opposed to a reduction in weight. However, since BMI is linearly related to weight, proportional changes in weight generate proportional changes in BMI. For example, consider a man who is 1.83 meters (6 foot) tall and weighs 100 kilograms. The corresponding BMI value is 30.[8] Therefore, a 5 per cent reduction in BMI will reduce the BMI value to 28.5. A 5 per cent reduction in body mass is equivalent to about a 5 kilogram weight loss. A 10 per cent reduction in BMI will reduce the BMI value to 27, which is equivalent to about a 10 kilogram weight loss. However, if this individual weighed 150 kilograms, his BMI value would be 45. In this case, a 10 per cent reduction to a BMI value of 40.5 would be equivalent to a weight loss of about 14 kilograms. This relationship indicates that reductions in BMI will be associated with a greater absolute reduction in weight loss, the greater the BMI value.

Another point worth noting is that the policy scenario estimates are for one year only. These figures do not take into account the potential lifetime savings associated with a reduction and maintenance of a healthy BMI value. Clearly, the above estimates indicate that if younger Australians (aged 20-39) were able to reduce and maintain their

BMI value within a 'healthy' or desirable BMI range, then the potential lifetime cost savings would also be considerable.

The policy scenarios have focussed on the benefits (cost savings) of weight reduction, and paid no attention to the methods available for generating weight reductions, and the costs of these. Yet information on the effectiveness of alternative policies, such as educational campaigns, advice from health professionals, subsidised diet food, and so on, is required before cost-effective interventions can be considered. Nevertheless, the estimates presented here give some guidance on how to plan such interventions. For example, it indicates the different magnitudes of benefits to be derived from targeting particular groups rather than the population as a whole. Effective and efficient policy needs to derive information on the costs of securing weight loss among these different groups: is it cheaper to generate an across-the-broad weight loss (of, say, 10 per cent) or a reduction in the number of obese people? The are the types of questions that need to be examined.

Although these policy scenarios have limitations, they nonetheless show that a moderate reduction in BMI could lead to substantial cost savings in the medical sector. Furthermore, these estimates could also be used to provide health policy recommendations such as the implementation of weight loss programs that are designed to target overweight and obese individuals. The final chapter of this thesis provides a summary of this research project, addresses study limitations, and puts forth areas for continuing economic research.

---

[8] BMI: $100/1.83^2 = 29.86$. The BMI values have been rounded to the nearest whole number.

CHAPTER 7: CONCLUSION

7.1 Introduction

The purpose of this research project was to provide answers to the following research questions posed in the introductory chapter. For convenience, these four questions are summarised below:

(1) Is there are casual relationship between BMI and RF?;

(2) Do these RF lead to increased MCU?;

(3) Does obesity have an independent impact on MCU; and

(4) To what degree might weight loss reduce the cost associated with MCU?

Given that the results and policy analysis has been reported in the previous chapters, it is now possible to summarise the key findings, address study limitations, and identify areas for continuing research.

7.2 Summary of Key Findings

The key findings presented in this thesis indicate that overall, there is a statistically significant positive association between MCU and the level of BMI. There is also strong evidence indicating that there is, in fact, a causal relationship between BMI and RF, and that RF lead to increased MCU, and hence costs. It is also worth noting that the introduction of the six RF variables significantly attenuated the independent impact of the BMI variables. In other words, obesity exerts its damaging influence primarily

through this particular set of risk factors. Overall, there is evidence to suggest that obesity in and of itself does not exert an independent impact.

The results reported in chapter 5 were used in the subsequent chapter devoted to policy implications. Poisson regression estimates were combined with unit cost data to generate MC estimates for various segments of the Australian population. Moreover, several policy scenarios were examined to provide an indication of the magnitude of the potential cost savings associated with a reduction in body size. Results from these policy scenarios indicate that the annual cost savings are substantial. For instance, if all obese individuals aged 20-64 with a BMI $\geq$ 30 reduced their BMI value by 5 per cent the annual cost savings would be about $158 million per annum. For a 10 per cent reduction in BMI, the annual cost savings would amount to about $310 million per annum. These findings indicate that public health policies designed to target obese individuals could result in considerable medical care utilisation cost savings.

7.3 Study Limitations

As with any study, there are always a number of limitations. In this study, MCU estimates were restricted to (i) the number of doctor visits, (ii) the number of other health care professional visits, and (iii) the number of hospital visits. Other resource costs such as laboratory tests and the level of prescription drug use were not included. As previously noted, this limitation would most likely bias downwards the potential cost savings estimates. In addition, there are also limitations associated with using short-term cross-sectional data to draw inference about long-term dynamic processes.

Another limitation with this project is that the data are self-reported. For this reason, data on variables such as height, weight, and health status can be open to a number of problems. For instance, there could be a problem with recall bias, which in turn, could lead to under-or over-reporting of certain items in the 1995 NHS. Also, it could be argued that there might be a tendency for respondents to over report their height and under report their weight since height is 'good' and weight is 'bad' in this society. This unfortunately is a limitation associated with self-reported data.

Furthermore, all the obesity-related risk factors such as diabetes mellitus, hypertension, and heart disease were also self-reported and some of these conditions may be misreported and, hence miscoded. Thus, this research project may have actually under estimated the magnitude of the association between MCU and the level of BMI. However, in the absence of large-sample 'objective' data, the 1995 NHS probably provides the best option for comprehensive research on these relationships.

7.4 Areas for Continuing Research

The research reported here is one of the first economic studies to use individual-level (i.e., microeconomic) data to model the association between MCU and the level of obesity (as defined by the body mass index). There are a number of potential economic research areas related to obesity warranting further investigation.

To begin with, it would be interesting to examine whether the relationships reported in this research project are similar for alternative datasets. Ideally, a longitudinal dataset would be extremely useful as the relationship between MCU and BMI could be tracked over time. Longitudinal data would also allow an investigation of the dynamics of

BMI; for example, whether weight gains tend to be continuous or tend to zero (i.e., a stable BMI). Furthermore, it would also be interesting to examine these relationships under alternative health care systems in other countries such as the United Kingdom and Canada.

Another possible research area would be to examine the relationships between lost productivity (such as work days lost) and the level of BMI. As indicated in the obesity cost-of-illness literature, indirect costs associated with obesity are considerable. Further analysis of the 1995 NHS data may allow exploration of this issue. For example, it would also be challenging to generate estimates of the intangible cost related to obesity. There is also evidence to suggest that there are psychological costs associated with being obese, especially for women. Subsequently, there could be potential research into the development and administration of instruments designed to measure the impact of obesity on an individual's quality of life. The key point to stress, however, is that there are a number of potential economic research areas in the field of obesity.

7.5 Concluding Comments

This research project attempted to answer a number of questions about the relationship between MCU and the level of obesity for a sample of the Australian population. Following an extensive review of the current literature a method was developed to quantify and examine these relationships. Specifically, data extracted from the 1995 NHS were used to model these relationships. Overall, the results supported the working hypotheses and the subsequent policy analyses indicated that potential cost savings associated with reduction in BMI could be substantial.