Adverse selection and risk selection in unregulated health insurance markets: Empirical evidence from South Africa's medical schemes

Thesis submitted to the University of London
in fulfillment of the requirements for the degree of Doctor of Philosophy

Birgit Hansl

Department of Social Policy
London School of Economic and Political Science
July 2004

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Health insurance arrangements developed in various social settings as a means of pooling health risks and health resources in order to protect members' income against unpredictable health costs but also in order to guarantee their access to health care. Problems of unregulated health insurance markets, like adverse selection and risk selection, are frequently discussed in academic and political circles in the context of either inefficiency or inequity. Though interest in regulation as a health sector reform instrument is growing, empirical studies of unregulated health insurance markets are still rare, particularly, in low and middle-income country settings.

This thesis contributes to the body of research and literature that attempts to identify empirical evidence for adverse selection and risk selection. It aims to examine the following research question: Are unregulated health insurance markets characterised by adverse selection and/or risk selection and do they thereby create inefficiency or inequity? The objective is to demonstrate empirically whether or not these markets experience selection processes. First, this thesis derives a group method for empirical investigations into adverse selection and risk selection from which testable hypotheses can be derived. Second, this method is applied to case study data from a middle-income country. Longitudinal panel data is analysed, describing South Africa's health insurance market of medical schemes in the context of its post-deregulation experience over the four-year period 1995-1998, after premia risk-rating was legalised.

The interpretation of the empirical results leads to three main findings. First, intense competition in the contested health insurance market causes favourable risk selection of low risks into and out of medical scheme plans. Second, unfavourable selection by medical scheme plans in the form of dumping high risks can be concluded. Third, there is no evidence for adverse selection and the typical adverse selection cycle cannot be observed. Exploring the policy relevance of the results, it is concluded that the effects of less health insurance regulation, in the context of middle-income country health sector reforms, conflicts with the common health policy objective of equity. More competition and efficiency comes at the price of less equity in health care access for the poor and sick, confirming the known efficiency-equity trade-off.

Acknowledgements

This thesis was made possible through the support of several research scholarships and grants. First, I would like to thank the London School of Economics for supporting me with a generous Research Student Scholarship from 1997/1998 to 2001/2002. I extend my acknowledgements specifically to the Department of Social Policy, which additionally awarded me the Tittmuss-Meinhard Research Scholarship in Social Medicine for the three successive academic years between 1998 and 2002. I also feel very honoured that the Gottlieb-Daimler & Karl-Benz-Stiftung (Germany) decided to sponsor this research within its Doctoral grant program (1999-2000) and I am very grateful for the excellent networking it provided among their group of Doctoral fellows.

I owe my academic development and passage to the completion of this thesis to my two formidable teachers and supervisors, Dr. Lilani Kumaranayake from the London School of Hygiene and Tropical Medicine and Prof. Julian Le Grand from the London School of Economics and Political Science. My supervisors not only provided me with academic guidance and prompt and constructive comments on my drafts, they also paid close attention to my progress and gave me moral support. Since English is not my mother tongue, I am in particular thankful for their untiring efforts to make this thesis more readable.

During my fieldwork in South Africa, I received immeasurable help from several people. Most of all I wish to thank Dr. Neil Söderlund from the Centre for Health Policy, University of Witwatersrand in Johannesburg, who familiarised me with the South African medical scheme market and introduced me to experts in the field. He also encouraged me to explore the data set used in this thesis after our initial research project expired. Special thanks are due to the former Registrar of the Medical Schemes, Danie Kolver, and to Ryno van Zijl for supplying additional information and considerable assistance in the interpretation of the studied medical scheme files at the Office of the Registrar in Pretoria.

My heartfelt thanks go to my family and friends, for their company, encouragement and patience on this sometimes seemingly endless and lonely journey. Most of all I thank Neil Pogorelsky, my best friend, determined partner and true companion, who always reminded me that love travels faster than light.

Abbreviations

AAPCC	Adjusted Average Per Capita Cost
ANC	African National Congress
AS	Adverse selection
AR(g)	Auto-regressive process of degree g
CI	Confidence interval
CLASS.	Classification
COEF.	Coefficient
CO2	Co-payments
CUM.	Cumulated
DEN	Dental benefits (per membership month and member)
DEPM	Dependents
DOH	Department of Health
E.G.	Exampli gratia (for example)
FREQ.	Frequency
GCIS	Government Communication and Information Systems
GDP	Gross Domestic Product
GEAR	Growth, employment and redistribution strategy
GEE	General estimating equation method
GEN(F)	Gender (female)
GLM	Generalised linear models
GNU	Government of National Unity
GP	General practitioner (benefits) (per membership month and member)
HMO	Health Maintenance Organisation
HR	High risks
HST	Health Systems Trust
ICGM	Income group
I.E.	Id est (that is)
ID	Identification
IFP	Inkatha Freedom Party
IMF	International Monetary Fund
JC	Joiner-control group identifier

LC	Leaver-control group identifier
LR	Low risks
MAR	Missing at random
MAX	Maximum
MCAR	Missing completely at random
MIN	Minimum
NGO	Non-governmental organisation
NP	National Party
N.S.	Not significant
OBS.	Observations
OPT	Optical benefits (per membership month and member)
OR	Odds ratio
PAR	Paramedical benefits (per membership month and member)
P or PROB	Probability
PERC	Percent
PMBS	Prescribed Minimum Benefits
PPO	Preferred Provider Organisation
PRV	Private hospital benefits (per membership month and member)
PUB	Public hospital benefits (per membership month and member)
R	Unstandardised residual
RA	Race, Asian
RB	Race, Black
RC	Race, Coloured
REF. CAT.	Reference category
R(S)	Standardised residual
RS	Risk selection
RW	Race, White
SC	Switcher-control group identifier
SHI	Social health insurance
SPE	Specialists' benefits (per membership month and member)
STD.DEV.	Standard deviation
STD.ERR.	Standard error
TAC	Total claims (per membership month and member)
TCO	Total contributions (per membership month and member)

UNECA	United Nation Economic Commission for Africa
URB	Urban residency
US	United States of America
WHO	World Health Organisation
W/OUT	Without

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Thesis topic and scope

This thesis' topic is adverse selection and risk selection in health insurance markets. It will address gaps in the literature, specifically in regard to empirical evidence for adverse selection and risk selection in low- and middle-income countries. Both selection processes are frequently discussed in academic and political circles in the context of either inefficiency or inequity. However, these processes are usually not empirically assessed, likely because they are hard to identify. This thesis is an attempt to rectify this omission.

More specifically the research has theoretical and practical aims. The theoretical aim is to derive a method for empirical investigations into adverse selection and risk selection that is easily applicable in low- and middle-income country settings. The practical aims are twofold. The first is to conduct an empirical analysis, which tests for adverse selection and risk selection. It will apply this method to data from South Africa's medical scheme market, thereby increasing understanding of the behaviour of health insurance market participants in the context of a middle-income country. The second practical aim is to provide policy-relevant information on how less regulation in middle-income country health sector reforms - in particular, unregulated health insurance arrangements - might impact the common health policy objectives of efficiency and equity.

The relaxation of regulation (or deregulation), as a type of privatisation and decentralisation in the context of health sector reforms, often aims to overcome the dilemma of poor public health care financing and provision (World Bank 1993). Health sector regulation became an area of interest as the role of the public health sector in low- and middle-income countries declined. At the same time the non-governmental sector,

both non-profit and for-profit, grew (Kumaranayake 1997). Regulation can be seen as a potential middle ground between "all or nothing approaches" of public versus private health care financing and provision. However, deregulation leaves these parts of the health care sector unregulated and *de facto* operating as private and for-profit. One example is the unregulated South African health insurance market of medical scheme plans, which became subject to a series of deregulations by the South African government between 1989-1994.

While deregulation in Central Europe and the United States was a rather selective movement, leaving areas such as health unscathed or even more regulated (Peltzman 1989), the picture was different for policies imposed on or recommended to low and middle-income countries. The structural adjustment programs of the International Monetary Fund (IMF) and World Bank explicitly emphasised deregulation strategies that would increase the role of the private health care sector as compared to the public sector. However, practice shows that the benefits associated with these reform instruments do not always materialise (World Bank 1998).[1] Nevertheless, critiques of this kind are largely based on anecdotal and not empirical evidence, which often makes them contested by opponents of more liberal viewpoints.

Despite the growing interest in deregulation as a health sector reform instrument, studies of unregulated markets are still rare. This gap is particularly noticeable for empirical investigations in the context of low- and middle-income countries. The reason for this is twofold. First, there is a general shortage of data for health insurance markets here. The data problem in these countries is more serious since data collection here is seldom a feasible priority. Second, although theory highlights adverse selection and risk selection as anticipated problems in unregulated health insurance markets, empirical evidence for selection processes is hard to obtain. This is because researchers face distinct methodological challenges of assigning observed behaviour outcomes to theoretically predicted motivations. Thus while there is a well-developed theory on these selection processes, it does not lend itself very well to empirical data testing.

[1] The outcome of a critical internal assessment of this strategy by the World Bank (1988): Adjustment Lending: An Evaluation of Ten Years of Experience, illustrates that structural adjustment programs undertaken by 15 sub-Saharan African countries failed in many areas. However, a look into the actual design of recent structural adjustment programs shows a high resistance to this critique. These programs still emphasise a greater role for NGOs and the private sector in the provision of health care (IMF 1997); the privatization of public enterprises in the health sector (IMF 1998a), and promoting and strengthening of the private sector (IMF 1998b).

This thesis will contribute to the body of research and literature that attempts to identify empirical evidence for adverse selection and risk selection in health insurance markets. The thesis will utilise a method for empirical investigations into adverse selection *and* risk selection from which testable hypotheses can be derived and will apply this method to data from South Africa's medical schemes for their post-deregulation period, 1994-1998. In doing so the thesis will contribute to the discussion of the particular problems that low- and middle-income countries face after their health insurance markets become unregulated. The conclusion will be that there is need for more, not less, regulation in health insurance markets in order to achieve more equity. The thesis will also discuss policy recommendations for the South African case in light of the results obtained from the empirical analysis.

1.2 Framework of the thesis

This thesis will be embedded in a framework that contains three parts: the research question, the main research objective and research aims, and the main research hypothesis and research subhypotheses. The purpose of the following sections is to introduce each of the three parts of the framework. Section one will introduce the research question and the case study context. Section two will state this thesis' research objective and derive two specific research aims. Section three will formulate the main research hypothesis and specify the three subhypotheses for this thesis' empirical investigation.

1.2.1 Thesis research question and case study context

This thesis' research question asks whether unregulated health insurance markets are characterised by adverse selection and/or risk selection, thereby creating inefficiency or inequity. Unregulated health insurance markets either emerge from purely unregulated health insurance arrangements or they emerge from initially government or state regulated arrangements that become subject to various degrees of deregulation at some point in time. As a case study, this thesis will analyse the unregulated health insurance market of a middle-income country - South Africa's main health insurance market of medical schemes, in the context of its post-deregulation experience over the four-year period 1995-1998, after premia risk-rating was legalised. Inefficiency is defined within this thesis as the observation that when demand and supply for unregulated health insurance

do not meet, the market equilibrium is inefficient, and, in the extreme, this market fails. Inequity is defined as inequitable access to unregulated health insurance. However, for the South African case study it will be shown that this also implies inequity in access to good health care.

Health insurance arrangements developed in various social settings as a means to pool health risks and health resources in order to protect the income of its members against unpredictable health care costs but also in order to guarantee their access to health care. Early health insurance arrangements were typically organised as mutual funds with quite restricted enrolment. They were based on employment or on small manageable groups of geographically concentrated members who were charged community-based premia. Community-based premia are identical contributions for all insured members, regardless of their individual risk. Community-rating implies cross-subsidisation (where high-risk individuals receive a subsidy from low-risk individuals thereby increasing their access to health insurance coverage). However, on the fund level premium contributions and paid-out benefits need to be balanced. As non-profit organisations, these mutual insurers were often legally recognised and subject to state regulation.

As health insurers grew, other, for-profit health insurance organisations developed, often in competition with the earlier mutual arrangements. These for-profit insurers usually operated under unrestricted enrolment. They were based on general population members, who wished to join and who were charged risk-based premia. Being competitive insurers, they did not only have to balance collected premium contributions and paid-out benefit, but had to generate profits. In order to maximise profits, these insurers charged risk-based premium contributions and competed for households with lower health benefits compared to their paid premium contribution, so called low risks.

Risk-based premia define different contributions for each insured member, depending on the individual risk. Households are referred to as low risks (as compared to high risks), if their paid-out health benefits lay below their premium contributions. High risks constitute households for whom benefits exceed contributions. Households might end up with benefits exceeding their premia for several reasons, for instance if elderly, if chronically sick or if having a large number of dependants they might be more likely to have episodes of ill-health that incur health related costs. Risk-based premiums and selected pooling of low risks imply less or no cross-subsidisation.[2]

[2] Van de Ven et al. (2000) refers, in the context of competitive health insurance markets, to two principles of equivalence and solidarity (or fairness). According to their paper the equivalence principle implies that an

Competitive insurers often practice risk selection in the form of 'cream skimming' and pull low risks out of mutual insurers' risk pools, while at the same time neglecting high risks by either dumping them out of their insurance or refusing their entry. As a result mutual insurers might not only experience a pull-out of low risks due to risk selection by the for-profit competitors, but might also experience a push-in of high risks into their more affordable community-rated coverage. The later trend can lead to adverse selection processes. The influx of more costly high risks leads to higher average premia that are unattractive for other, lower risks, initiating their exit. Both described processes of adverse selection and risk selection dramatically worsen mutual insurers' risk pools and are likely to cause those pools to cease to exist. Their ex-clientele of particularly high risks cannot afford private, risk-related health insurance coverage. Since there is demand for mutual insurance coverage, but it cannot be supplied under competitive market conditions, this market result is not efficient. Most of all this market failure raises equity concerns, since access to health insurance and health care for the poor and sick is blocked. Other arrangements, like public medical programs have to fill that gap at considerable social costs.

A similar market outcome occurs if legislators deregulate mutual insurers in order to create equal market conditions as compared to unregulated, for-profit insurers. Providing similar conditions through deregulation typically means allowing risk-rated premia and free enrolment. Without the cross-subsidisation of community-rated premia, high-risk households most in need of insurance coverage would be no longer able to afford insurance and would have no health care access. Also, mutual insurers might become subject to adverse selection in the short-run if they fail to enter the competition for low risks by applying strict risk selection strategies. As a result of adverse selection mutual insurers might go bankrupt. Thus it can be argued that as a result of health insurance deregulation the envisioned positive effects of more competition and efficiency come at a price of more inequity.

This thesis' case study of the South African health insurance market of the early 1990s is one example where a legislature deregulated previously regulated mutual insurers in order to improve their competitive position against existing for-profit insurers. The general problem of this type of health insurance deregulation and its particular application to the case of South Africa can be described as follows. In the 1990s two

insurer has to break even on each insurance contract, while the solidarity principle stresses cross-subsidisation. Their focus of discussion is the incompatibility of the equivalence principle and the solidarity

distinct health insurance arrangements, subject to different regulations, were operating side by side in South Africa. The health insurance market of concern for this thesis is the medical scheme market that has its historic origins in mutually organised, employment-based insurance arrangements. This part of the market was subject to strong government regulation since the 1960s that recognised them as non-profit entities and prevented any risk-related premia-rating (Medical Scheme Act 1967).[3] Medical schemes cover 20% of the South African population, while the other, much smaller part of the South African health insurance market, for-profit health insurers, covers 4% of the population. Only members of medical schemes or other private health insurers have access to good quality health care provided in the private sector. This sheds light on another equity dimension of health insurance access in South Africa: the majority of South Africans have no access to good quality health care because they are uninsured. Also, if formerly insured South Africans become subject to unfavourable risk selection they will loose not only any access to good health care, but will also add an additional burden to the modest resources of the public sector for which the remaining two-thirds of South Africans compete.

For-profit health insurers emerged much later than the medical scheme market and were at no point subject to government regulation that would prevent premia risk-rating. By the late 1980s competition between the two health insurance arrangements grew and, at the same time, medical costs escalated. However, medical schemes could not effectively compete with for-profit insurers because medical schemes were required to apply community-rating while for-profit insurers could risk-rate premia.

The South African government deregulated the medical scheme market by the early 1990s (Health Systems Trust 1996, 1998, 2003). In the 1989 amendment to the Medical Scheme Act, rules disallowing risk-rating of premia and the stipulation of minimum reimbursement rates were removed from the statute books. A further deregulation in 1994 did away with the regulation of reimbursement rates altogether, and allowed contracting and vertical integration between providers and medical schemes, paving the way for managed care arrangements. Risk-rating of premia was expected to create fairer market conditions for the competition between medical schemes and for-profit insurers.

Thus the prior to 1989 *regulated* medical schemes became *unregulated* because of *deregulation* by the South African government between 1989 and 1994 (Health Systems

principle in competitive health insurance markets.

[3] The Act also established two bodies, the Council for Medical Schemes and the Registrar of Medical Schemes to fulfil the executive functions of the scheme, i.e. their regulation. Medical schemes have to submit statutory returns to the Registrar of the Medical Schemes on an annual basis (Rama and McLeod 2001).

Trust 1996 and 1998, Soderlund and Hansl 2000).[4] The data, which will be used in the thesis, covers the four-year period between 1995-1998. Some medical schemes might have adjusted to the deregulation of 1989 earlier than 1995 when the observed data starts, so the observations for this particular unregulated market cannot be proven to be results of the regulatory change.

However, this thesis' case study will describe a post-deregulation experience of a health insurance market. This is based on the observation that deregulation rules, which allowed risk-rating, were only fully implemented by the medical schemes in the mid-1990s. This conclusion is drawn from previous research (Soderlund and Hansl 1999), including interviews with the Registrar of Medical schemes, Danie Kolver, and his assistant Ryno van Zijl, and a document review of medical scheme files for the years between 1989 and 1998 at the Registrar's office in Pretoria. These sources verified that schemes did not start applying risk-adjustment strategies that included premia risk-rating until 1993-1995. In fact very few schemes started any risk-adjustment strategies by 1993, and most schemes took at least one or two more years before they established strategies that included premia risk-rating. Thus while a before-and-after deregulation scenario for the analysis of the thesis' case study cannot be performed with the available data starting in 1995, it can be established that the analysis of the unregulated South African medical scheme market accurately describes a post-deregulation experience.

1.2.2 Thesis research objective and aims

This thesis' research objective is to empirically analyse whether unregulated health insurance markets experience adverse selection and risk selection. There are few empirical studies that explore these selection processes in health insurance markets of low- or middle-income countries. The limited number of publications can be explained in part by the shortage of good data in low- and middle-income countries. But more importantly, distinct methodological challenges have likely inhibited this research, because the well-developed theory on these selection processes does not lend itself very well to empirical data testing. Thus in order to be able to conduct an empirical investigation into evidence

[4] When referring to the regulation of the medical scheme market prior to 1989 this will refer to the mandatory rule of community-rated premia for medical schemes. Regulated medical schemes then are medical schemes under mandatory community-rating requirements. Deregulation of the medical scheme market will refer, in particular, to the legalisation that introduced risk-rated premia by the beginning of the 1990s. Due to this deregulation medical schemes became unregulated. The unregulated medical scheme market will refer to the medical schemes after they actually implemented risk-rated premia.

for adverse selection and risk selection, the following two specific research aims have to be fulfilled:

1. Derive a method applicable to empirical investigations into adverse selection *and* risk selection with available quantitative techniques and easily accessible data.
2. Apply the method to an empirical analysis of case study data of an unregulated health insurance market in a middle-income country: panel data of insured households in South Africa's medical scheme plans, 1995-1998.

The case of the South African medical scheme plans provides a natural experiment for studying adverse selection and risk selection in unregulated health insurance markets.[5] Medical schemes are either self-administered (by their own staff) or professionally administered by for-profit companies (Soderlund and Hansl 1999 and Rama and McLeod 2001). The data for the thesis' empirical analysis originated from a for-profit holding company in South Africa that manages the administration of several medical insurance schemes.[6] The available medical scheme data contains records from insured households in the form of a panel.[7] This panel consists of 1,011,735 observations of 353,458 insured households over a four-year observation period between 1995 and 1998. The described insured households are members of 29 medical schemes. Medical schemes might have several plan options; however, some have only one plan. Altogether the 29 schemes in this sample have 49 plan options. The 29 schemes represent about one fifth of South Africa's registered medical schemes between 1995-1998.[8]

Each of the insured households has observations for at least one year and for a maximum of four years. The observations contain information on household insurance status (i.e. on the beginning and, if applicable, end of coverage), on the principal members' socio-economic characteristics, such as age and income, and on households' chosen coverage and utilisation, such as the plan name, premium contribution and actual claims and benefits.

[5] The document review of previous research at the Registrar' of the Medical Schemes office showed that between 1993 and 1995 medical schemes implemented risk-management strategies in three forms. First, benefit ceilings, second, risk-related/age-related premia, and third options within the insurance schemes with different levels of coverage, contributions, co-payments and deductibles. Only a few schemes introduced the legalised managed care options and saving accounts. However, this does not establish empirical evidence for selection processes. Also, risk selection strategies can be much more sophisticated than explicit contractual discrimination (like age-rated premia).

[6] The data was provided by this holding company proceeding a research project conducted by the author on regulatory aspects of the South African health insurance business in 1998.

[7] Panel data are also called longitudinal data. They follow a given sample of subjects over time (providing multiple observations on each subject in the sample).

[8] There were altogether 160-180 registered medical schemes between 1995-1998 (Soderlund and Hansl 2000, Rama and McLeod 2001).

1.2.3 Thesis research hypotheses

The main research hypothesis for the thesis' empirical investigation is that South Africa's unregulated medical schemes encountered a lack of market-wide risk pooling and a lack of risk pooling within the market (between insurance plan options). This is based on the assumption that unregulated health insurance markets do not pool risks equally, since it is likely that adverse selection and risk selection occur (Soderlund and Hansl 2000). Adverse selection and risk selection might occur as a result of market competition and the supply of health insurance choices.

Health insurance choice with different insured individuals selecting different plans to meet different individual preferences. It is not only beneficial to have these preferences reflected in market outcomes; health insurance choice also promotes efficiency (Cutler and Zeckhauser 2000). However, the cost of different individuals typically varies substantially, depending on the insured's risk. As Cutler and Zeckhauser (2000:607) point out: "Whom one pools with in the health insurance dramatically affects what one has to pay." If the market can efficiently sort high and low risks into either more generous or less generous plans, while charging individuals their expected costs, a separating equilibrium can exist (Rothschild and Stiglitz 1976). However, even if efficiency could be obtained - here in form of a separating equilibrium - charging high risks substantially more is widely seen as inequitable.

If plans can only charge average prices and/or cannot distinguish between different insured, risks adverse selection occurs. Typically the more generous insurance plans experience a disproportionate number of high risks, so that their risk pools worsen (Nyman 2003). Insurance plans also have an incentive to practice risk selection by plan manipulation, i.e. distorting plan provisions and benefits to attract low risks and repel high risks. Risk selection results in some insurance plans disproportionately enrolling low risks, while other are left with a disproportionate number of high risks (Newhouse 1997).

Thus adverse selection and risk selection can change the risk pools of individual insurance plans and/or the whole market risk pool. The main research hypothesis expects that one can observe a lack of risk pooling between high and low risk enrolees in South Africa's medical schemes. This does not exclude the possibility of a separating equilibrium, which might be indeed efficient but yet not equitable. Although the thesis's data and the empirical analysis design do not lend themselves to an explicit test for the existence of a separating equilibrium this interpretation of the results will be discussed.

South Africa's deregulation, which allowed medical schemes to implement risk-management strategies, transformed formerly highly regulated non-profit medical schemes into quasi-private for-profit schemes. Since then several reports indicated unbalanced risk concentration in the risk pool of the medical schemes in general and in some particular medical scheme plans, which could be attributed to a variety of factors including adverse selection and risk selection (Health Systems Trust 1998, Soderlund and Hansl 1999, Hansl and Soderlund 1999, Soderlund and Hansl 2000).[9] Three subhypotheses for this thesis' empirical investigations into South Africa's unregulated medical scheme market were formulated:

Hypothesis 1

There is competition for low risks between medical schemes and for-profit insurers. The previously regulated (observed) medical scheme market segment is contested, and experiences a disproportionate loss of low risks due to:

A. Risk selection by (external) for-profit insurers for low risks in the medical scheme market segment, and (initiating)
B. Adverse selection with low risks exiting medical schemes, and, in particular, exiting high-risk plan options because insurance coverage is comparably expensive and they can find better coverage at lower costs elsewhere.

Hypothesis 2

In the competition medical schemes follow several risk-management strategies to protect their scheme and plan risk pools including:

A. Risk selection that discourages the entry of high-risk insurance applicants;
B. Risk selection that attracts the entry of low risks with self-selection low-risk plans;
C. Risk selection that prevents the exit of low risks by offering low risks to switch to attractive low-risk plan options;
D. Risk selection in the form of pressuring high risks to switch into high-risk plan options; and
E. Risk selection by discouraging high risks to continue medical scheme coverage, particularly if they belong to high-risk plans.

Hypothesis 3

High risks, who are either seeking medical scheme coverage or who are already insured in the previously regulated medical scheme market follow strategies of:

[9] For example, increasing benefits and utilisation might be partially attributable to a changing demographic profile of medical scheme plan members but adverse selection seems a very likely explanation too.

A. Adverse selection by choosing to enter medical scheme plan coverage, in particular expensive and comprehensive high-risk plan options; and
B. Adverse selection by choosing to switch into medical scheme plans with expensive and comprehensive coverage, typically high-risk plan options.

The formulation of these three research hypotheses is due to the peculiarities of the South African health insurance market, where, in addition to the medical schemes, a small number of for-profit insurers operated. These for-profit insurers were in competition with the medical schemes for attracting low risks. However, only after the deregulation legalised risk management strategies also for the medical schemes were they able to compete seriously with the for-profit insurers. The competition for profit-maximising low risks naturally comes in the form of risk selection. However, since community-rating is widely abolished, it might be more important for high risks to strategically select appropriate insurance coverage, as long as they can afford it. If medical scheme plans were successful in applying risk selection strategies one can expect to observe a better overall risk pool or some better individual plan risk pools. However, if the for-profit insurers proved to be better at attracting low risks, or if high risks could successfully adversely select themselves into medical scheme plans, one can expect to observe a worse overall medical scheme risk pool or some worse-off individual plan risk pools.

Table 1-1 on the following page summarises the framework of the thesis by listing the research question, research objective and aims and research hypotheses.

1.3 Organisation of the thesis

This thesis is divided into seven chapters (including this introduction). The next chapter reviews the analytical context and the theoretical concepts of adverse selection and risk selection. This will provide the rationale for this thesis' research question. First, the economic model of human behaviour will be presented as the analytical context of the thesis. It will then review the literature regarding adverse selection and risk selection in health insurance markets, highlighting key theoretical concepts. A critical review of empirical studies, which have attempted to conclude the existence of evidence for adverse selection and risk selection utilising various methodological approaches, will be conducted. In this review gaps in the literature will be identified, which this thesis aims to address. Finally, based on the identified gaps in the literature, chapter two will develop the thesis research question and research objective.

Table 1-1: Thesis framework

RESEARCH QUESTION
Are unregulated health insurance markets characterised by adverse selection and/or risk selection thereby creating inefficiency or inequity? Case study: South Africa's unregulated medical scheme plans in the context of their post-deregulation experience after premium risk-rating was legalised.
RESEARCH OBJECTIVE and AIMS
Empirical analysis of whether unregulated health insurance markets experience adverse selection and/or risk selection. Aim 1: Derive a method applicable to empirical investigations into adverse selection *and* risk selection with available quantitative techniques and easily accessible data. Aim 2: Apply this method to an empirical analysis of case study data of an unregulated health insurance market in a middle-income country: panel data from insured households in South Africa's medical scheme plans 1995-1998.
RESEARCH HYPOTHESES
Case study: South Africa's unregulated medical schemes encountered a lack of market-wide risk pooling and a lack of risk pooling within the market (between insurance plan options). Hypothesis 1: There is competition for low risks between medical schemes and for-profit insurers. The previously regulated (observed) medical scheme market segment is contested, and experiences a disproportionate loss of low risks due to: A. Risk selection by (external) for-profit insurers for low risks in the medical scheme market segment, and (initiating) B. Adverse selection with low risks exiting medical schemes, and, in particular, exiting their high-risk plan options because insurance coverage here is comparably expensive and they can find better coverage at lower costs elsewhere. Hypothesis 2: In this competition medical schemes follow several risk-management strategies to protect their scheme and plan risk pools including: A. Risk selection that discourages the entry of high-risk insurance applicants; B. Risk selection that attracts the entry of low risks with self-selection low-risk plans; C. Risk selection that prevents the exit of low risks by offering low risks to switch into attractive low-risk plan options; D. Risk selection in the form of pressuring high risks to switch into high-risk plan options; and E. Risk selection by discouraging high risks to continue their medical scheme coverage, particularly if they belong to high-risk plans. Hypothesis 3: High risks, who are either seeking medical scheme coverage or who are already insured in the previously regulated medical scheme market follow strategies of: A. Adverse selection by choosing to enter medical scheme plan coverage, particularly expensive and comprehensive high-risk plan options; and B. Adverse selection by choosing to switch into medical scheme plans with expensive and comprehensive coverage, typically high-risk plan options

Chapter three is the first results chapter. It derives the method for this thesis's empirical investigations into selection processes in health insurance markets. There are three main sections in this chapter. In the first one the standard concepts of adverse selection and risk selection, usually explained using the time dimension, will be reformulated as grouping phenomena. Based on this a group method will be described from which empirically testable hypotheses can be derived. The second section will discuss the empirical strategy

that applies the group approach in regard to applicable quantitative techniques and data requirements. The third section will develop the research hypotheses.

Chapter four describes the conducted empirical analysis of adverse selection and risk selection, applying the methodology derived in the earlier chapter to data from South Africa's unregulated medical scheme plans. First, the data and sample derivation will be described. The data, in the form of a panel, describes several hundred thousand households insured in 49 medical scheme plans between 1995-1998. It contains information on households' socio-economic characteristics, plan membership and contribution, and claim and benefit levels. Applying the derived method, the data are divided into subgroups, which are described. Second, the empirical analysis strategy with logit models for panel data will be presented. The logistic regression analysis will apply marginal or population averaged models. Applying a confirmatory model building strategy three logit models are specified. The main research hypothesis is formulated into several subhypotheses for the South African case study setting and variable expectations are derived. Finally, the four-step model building process is described.

Chapter five will present the results of the empirical investigation and will interpret these results in regard to evidence for adverse selection and risk selection. First, the model estimation results for each of the three models will be presented successively. Second, the results for the model fit assessment, comprising several standard diagnostic statistics and methods, will be discussed. The interpretation of the model estimation results will follow. Finally, the results will be summarised and reviewed, particularly in regard to the previously formulated hypotheses. This will make it possible to either verify or falsify the main research hypothesis.

Chapter six is the discussion chapter and will focus on the discussion of three aspects. The first part will discuss the thesis' first contribution, the derived method for empirical investigations of adverse selection *and* risk selection. The second part will discuss selected thesis limitations, which are mostly related to the data and the empirical analysis. Finally, the thesis' policy relevance will be discussed - both in general terms and, in particular, for the South African case study. While the main focus here is on deriving policy implications from the analysis results, the key findings are also discussed in order to answer the originally posed research question.

The final chapter, seven, will pull together the knowledge and experience gained from the previous chapters. It will summarise the principal findings of this thesis, its

limitations and contributions to knowledge. A short outlook for future research will be presented.

Chapter 2

Analytical context & review of adverse selection and risk selection

2.1 Introduction

The primary purpose of this chapter is to provide the rationale for the thesis' research question. First, the chapter will define the analytical context of this thesis, with its underlying behavioural assumptions. Second, the literature regarding adverse selection and risk selection in health insurance markets will be reviewed. In the course of this review definitions and key concepts related to the research question, such as uncertainty, risk, and risk pooling in health insurance will be provided. This review will not only consider theoretical aspects of adverse selection and risk selection, but in particular, studies that have attempted empirical investigations into these selection processes. The goal is to identify gaps in the literature and to critically review different methodological approaches used in empirical studies. Finally, this chapter will develop this thesis' research question and research objectives based on the gaps identified in the literature.

After this introduction, part two will define the analytical context for this thesis, the economic model of human behaviour, which lends itself very well to the attempted analysis. In the literature adverse selection and risk selection are almost exclusively discussed under the behavioural assumption of the economic behaviour model. Thus the choice of it as the analytical context of this thesis' research is natural. The behavioural model is preferably used, because it can be applied consistently to all areas of human life,

including individual participation in health insurance arrangements. It also recognises the information problem and does not require that individuals be always perfectly informed.

The third part will review the theory of, and empirical evidence for, adverse selection and risk selection in health insurance. The literature focuses on adverse selection and market efficiency problems rather than risk selection and market equity problems. Generally, existing literature can be categorised into three types. The first focuses on abstract equilibria modelling considering selection problems and other theoretical issues of selection. Equally extensive, another type of published literature discusses selection problems within specific markets while often focusing on policy implications. Finally, a comparably small proportion of publications is dedicated to the empirical identification of selection processes using case study data. This thesis' literature review will recognise the theoretical concepts and definitions of adverse selection and risk selection and will discuss the effects of selection processes on efficiency and equity. However, the focus will be on reviewing the literature in regard to applied methodological approaches for the empirical identification of adverse selection and risk selection. The literature review covers publications on selection processes in the health sector and, in particular, health insurance. Where available, literature on low- and middle-income countries is preferably considered. Gaps in the literature will be identified. These gaps include the lack of studies on the impact of selection processes in the context of low- and middle-income countries and the particular lack of empirical studies that identify selection processes.

Finally, the fourth part will develop the thesis' research question and research objective, based on the identified literature gaps. The research question will ask: Are unregulated health insurance markets characterised by adverse selection and/or risk selection, thereby creating inefficiency or inequity? The research objective will be to empirically analyse adverse selection and risk selection in one unregulated health insurance market.

2.2 Analytical context: the economic model of human behaviour

This thesis performs an *economic* analysis of adverse selection and risk selection in unregulated health insurance markets. The economic model of human behaviour provides the analytical context for this analysis. The purpose of this section is to present the key assumption of the economic model of human behaviour. First, these behavioural

assumptions will be presented. Second, the role of information in the economic model of behaviour will be discussed.

2.2.1 Assumptions of the economic model of human behaviour

The primary assumption of the economic model of human behaviour is that behaviour is consistent in every function of human life and consists of rational self-interest.[10] This does not exclude other value-driven behaviour, but defines economically rational behaviour as the most influential. Since Becker's (1976): *Economic approach to human behavior* one can speak about a general economic model of behaviour with the following four assumptions (Blankart 1998, Kirchgaessner 1991, Frey 1980):

(1) methodological individualism declares the individual to be the only important acting unit. Every act is an individual act.[11] From an economic perspective only individuals can have interests.

(2) self-interest assumption assumes that individuals decide and behave in favour of their own advantage. Altruistic as well as vicious behaviour is generally not excluded, but it is assumed that, on average, egoistic and utility maximising behaviour is dominant.[12] A representative and responsible human being acts self-interestedly.

(3) given preferences and the change of restrictions assumes that (in economic analysis) preferences are taken as given and that restrictions are variable. If individual behaviour is changing than the reason for this is a change in the restrictions and not a change in

[10] This approach uses a consistent approach to human behaviour and does not split the human being into the *homo oeconomicus,* acting self-interested in an economic market context and a *homo politicus* acting according higher aims (for instance the 'public interest') in matters of the state. The *homo politicus* was a normative characterisation of human behaviour. Based on this characteristic, traditional economic and political science theories on regulation searched for the 'optimal state' by optimising a welfare function valid for *all* citizens. This reflects a normative system and a political system without personal decisions and responsibilities. Popper in his *Open society and its enemies* (1945) calls this a closed (tribal and collectivist) society with an organic or biological theory of the state. His concept of the open society describes people how they are and not as they should be. It confronts individuals with personal decisions and responsibility. This means a new individualism arises and the organic character of the 'closed society' (i.e. spiritual, physical or biological bonds in personal relationships) is weakened. The open society and its personal relations function more on abstract relations such as exchange and co-operation. "It is the analysis of these abstract relations with which modern social theory, such as economic theory, is mainly concerned." (Popper 1945:175)

[11] A state does not act as a whole or an organ, but on the strength of decisions of one or more individuals. With this assumption of methodological individualism the economic model of behaviour differs from any organic and functional view of the state, where the possibility of 'the state acts' is given. Thus phrases like 'public interest' have no room in this individualistic theory. However, they are a part of collectivist theories, which place the whole above the individual.

[12] This does not mean that egoism is necessarily a desirable state, but that it is a happy (and realistic) medium between a human being acting only altruistically or only viciously.

preferences.[13] The individual rationally adapts to the circumstances and tries to use the whole range of possibilities in order to maximise utility (assumption 2).[14]

(4) existence of relevant alternatives assumes that there are always enough alternatives between which an individual can choose. Substitution as a result of changing restrictions is possible. But substitution causes costs, which, again, are taken into account by an individual's economic calculation (assumptions 2 and 3).[15]

2.2.2 Information in the economic behaviour model

Individual economic decisions are driven by information and are influenced by the availability of this information. The model of economic behaviour is not based on perfect information. However, imperfect information does not mean that an individual is acting irrationally or unpredictably. Rather, the assumption is that the accumulation of information is only possible with certain costs. Rational individuals will only collect additional information as long as their expected marginal utility is greater than or equal to their marginal costs of collection. But this information equilibrium rests at a different level for different participants in a market, and particularly for buyers and sellers. For this reason the phenomenon of asymmetric information occurs.

The classical example is that of the competitive market for health insurance without regulation. Here the buyer has more information about his/her health risk than the seller: the costs of information gathering are lower for the buyer than for the seller. At a given premium those who think that their own costs of illness will lie above the price of the premium will purchase insurance.[16] Subsequently, the insurance company will realise

[13] Individuals behave differently because their circumstances change and not because their preferences shift. Although the later case is possible, it is seen as an exception because preferences develop in the long run, during processes of socialisation, and are therefore more stable than circumstances. Circumstances are variable in the short run and often dependent on fashions.

[14] Self-interested behaviour does not satisfy all wishes. Everybody maximises utility under the assumptions of supply, prices, income, rules and prohibitions etc.

[15] In general it is assumed that costs of substitution are in the long run lower than in the short run. That is why substitution often takes long. This does not change the fact that substitution is in general possible. Substitution has a stabilising effect on social processes: better alternatives are chosen, worse are given up. This last assumption makes the theory of the economic rational behaviour distinct to other theories in social sciences, especially the Marxist view that denies possibilities of substitution and predicts exploding social processes.

[16] Individuals decide to take out health insurance when they feel that they should hedge their income against the risk of health care costs. This decision is based on information of several kinds. First, it is related, within the broader context of uncertainty and risk, to the individuals' degree of risk aversion, which can range from not liking to take risks to enjoying taking risks. So far we have assumed that risk is unpredictable and illness is a stochastic event. However, subjects can also assess to a certain degree their risk for ill-health based on information or knowledge about their general health state and histories of illnesses. Furthermore, risk for ill-health can, to a certain degree, be influenced by behaviour like eating habits, alcohol consumption, exercise

that this given price is not covering costs. The insurance company will face adverse selection by high risks. The insurer will raise premiums with the result that the best risks will leave their contracts. For these individuals the price now lies above the risk costs. The high risks will stay and in the next round the costs will again exceed the revenues and the premiums will have to rise. This process may proceed until the market disappears.

In summary, one can derive three major conclusions from this example: first, the model of economic behaviour is valid under conditions of imperfect information, including asymmetry in information. Incomplete information yields different predictions than complete information, but both yield testable hypotheses. Second, individual rational behaviour can produce collective irrationality: despite the fact that there might be insurance companies who would offer insurance contracts and that individuals might agree that insurance is valuable, a market would not emerge under pure market conditions. Third, because this situation is unsatisfactory for all participants, it is predictable that individuals have an incentive to protect the health insurance system with other institutions than the market, e.g. state regulation. The literature discusses adverse selection (and risk selection) almost exclusively under the behavioural assumption of the economic model of human behaviour. This literature will be discussed in the next section.

2.3 Review of adverse selection and risk selection in health insurance: theory and empirical evidence

This thesis' research question concerns *adverse selection* and *risk selection* in *unregulated health insurance* markets. The key assumption is that unregulated health insurance markets face problems, in particular adverse selection and risk selection, which might be resolved with government market intervention. The purpose of this section is to justify the research question that links health insurance with the problems of adverse selection and risk selection. This will be done by reviewing the theoretical and empirical literature on the two selection processes in health insurance markets. Gaps in the literature that are specifically related to the empirical evidence for adverse selection and risk selection in health insurance will be identified. This section will first discuss uncertainty and risk, which provide the rationale for risk pooling in health insurance. Second, problems in

etc. The last two types of information introduce a deterministic component into uncertainty and risk (aversion). However, assuming a risk neutral individual, one could then predict that this individual would buy an insurance policy, as long its corresponding price would be less or equal to the expected health risk, depending on the budget (i.e. income) constraint of the individual.

unregulated health insurance markets, namely, adverse selection and risk selection will be discussed. Finally, existing empirical evidence regarding adverse selection and risk selection will be assessed.

2.3.1 Health insurance: risk pooling as a response to uncertainty

Health insurance developed as a means to reduce uncertainty of ill-health and its impact on a household's economic situation. Insurance can be seen as an effort to restrain the scope of uncertainty and to define its range (Padoa 1984). The negative effects that arise due to incalculable uncertainty might be restrained by the pooling of calculable risk. Thus risk pooling in insurance arrangements, not only health insurance arrangements, is the response to uncertainty.

Health insurance is based on statistical techniques that can calculate, and thus foresee, the probability of events of ill-health, which are going to take place within a given insured population. Although this statistical truth is generally reliable, health insurers clearly cannot forecast the events of ill-health for one specific individual. However, risk calculations and risk assessments performed by insurers are a bridge between the collective statistical certainty and the individual uncertainty.

A. Uncertainty and risk

The understanding of uncertainty in economics varied widely over the last century and often the difference between uncertainty and risk was not distinguished. At the end of the nineteenth century issues of change, risk and uncertainty became an explicit object of economic theorising by writers in the field of profit theory. Since changes are facts in every entrepreneurial activity, scientists were interested in a satisfactory explanation of profit and loss and its relation to uncertainty and risk.

Clark's dynamic theory of profit focuses on the entrepreneurial incentive of potential profit brought by innovation-induced changes in production methods (Clark 1893). He overlooks, in this way, the distinction between foreseeable and unforeseeable changes and that only the latter make profits possible (Wubben 1993). Hawley's theory of profit defines the assumption of risk or risk-taking as the distinguishing feature of an entrepreneur that will reward him with an uncertain amount of profit (Hawley 1901). He does not distinguish between indeterminate uncertainty and determinate risk and profit.

Haynes (1895) tries to combine approaches of dynamic profit theory and risk by differentiating between static risks present in a society and dynamic risks occurring due to

changes. He argues that statistic probability (under the assumption of homogeneity in outcomes) shows that perceived uncertainties and risks of an insurance holder are certainties for the insurer (Haynes 1895). Schumpeter's dynamic theory of profit divides risks into foreseeable and unforeseeable risks (Schumpeter 1921). Foreseeable risks are normally incorporated into economic plans by applying a range of methods for forseeable risk-spreading: incurring costs to guard against risks, risk avoidance, and cost accounting (including risk premia). However, his theory of economic development does not define uncertainty properly and does not distinguish it from risk (Wubben 1993).

Knight (1921) uncovered the negligence of uncertainty in economics and pointed out the confusion between risk and uncertainty: "Uncertainty must be taken in a sense radically distinct from the familiar notion of risk, from which it has never been properly separated."(Knight 1921:19). For Knight uncertainty is an aspect of the open future and an important characteristic of decision-making. Profit is the consequence of change, the effect of our imperfect knowledge of the future (Knight 1921), *and* the result of immeasurable risks. Thus, one can investigate the economic consequences of uncertainty but one cannot predict the underlying changes (Wubben 1993). According to Knight (1921) uncertainty reflects the immeasurability of outcomes in incomparable situations (violating the statistic probability assumption of homogeneity in outcomes). In economic life uncertainties are: (1) errors in predicting the future, and (2) adjustments made to future conditions (Knight 1921).

Uncertainty can be the result of indeterminacy or ignorance. Ignorance refers to a concept of subjective uncertainty, whereas the terms indeterminacy and unpredictability refer to the very notion of objective uncertainty (that is for at least a given time period and not explicitly related to a person and their opinion). Tversky and Kahnemann (1982) qualified uncertainty further with the adjective *external* when uncertainty can be attributed to someone's environment. *Internal* uncertainty may be attributed to someone's state of mind (Tversky and Kahnemann 1982). According to Wubben (1993) there exists additionally, the distinction between *endogenous* and *exogenous* uncertainty. Uncertainty is an inherent characteristic of the situation in the case of *endogenous* uncertainty only (Wubben 1993). Thus, coincidences can be classified as *exogenous, external* uncertainty, while ignorance is an example of *exogenous, internal* uncertainty. The purchase of health insurance coverage changes a subject's relation to a situation and also changes the situation itself. This would be a case if *endogenous, internal* uncertainty.

Risk is separated from uncertainty by measurability. Risk is based either on *a priori* probability distributions or empirically based probability distributions. In a real economic context the calculation of situations according to general *a priori* principles or empirically based probabilities is highly unlikely (Knight 1921), because the underlying situations of uncertainty are unique. Thus, it might be risky to buy or to sell health insurance contracts, but it is genuinely uncertain whether a specific insured person will be seriously ill or whether a specific insurance company will still be operating in the market in ten years.[17] Table 2-1 summarises the crucial distinction between uncertainty and risk.

Table 2-1: Attributes of uncertainty and risk

Attributes	Uncertainty	Risk
Measurability	No	Yes
A priori probability distribution	No	Yes
Empirically based probabilities	No	Yes

B. Risk pooling in health insurance

Health insurance arrangements essentially pool members' risk and resources in order to protect them from the uncertainty of unpredictable health costs that would limit their health care access (Soderlund and Hansl 2000).

The term risk pooling is used in various contexts within different research disciplines, depending on the risk that is to be pooled.[18] It refers generally to the accumulation of entities of different risks in one cluster in order to spread risk over a wide sample of these entities of different risks. These entities can range from abstract entities, like strategies or behaviours, to concrete entities like households or specific actions. In the context of health insurance, risk pooling is used in two different ways (n/e/r/a 1993, n/e/r/a 1996):

(1) Risk pooling refers to the pool of financial resources, i.e. individual member premia, which insurance providers accumulate and set aside, according to the terms of contract. These pooled financial resources are used throughout the contract period to pay or reimburse the health care costs of insured members in cases of ill-health. The amount of money remaining in the risk pool at the end of the contract (year) remains with the insurance provider as a profit, is rolled over, invested, allocated to health care providers, or paid-out to insurance members, depending on the insurance agreement.

[17] This leads to the conclusion that if scientists incorporate uncertainty and changes in their theories as if following a known pattern, they do not talk about uncertainty but fully calculable risks: "...the existence of a problem of knowledge depends on the future being different from the past, while the possibility of the solution of the problem depends on the future being like the past." (Knight 1921:313)

(2) Risk pooling also refers to the pool of an expected insured population, usually within a defined geographic location. Socio-economic characteristics of this population and their retrospective or anticipated health care utilisation can be used to determine members' health risk, expected revenues & expenses and thus anticipated claims liability.

Risk pooling is often specifically used in the sense of risk-spreading, and cross-subsidisation, implying that the high-risk individuals receive a subsidy from the low-risk individuals that increases their access to health insurance coverage (Van de Ven et al. 2000).

In summary, the presence of uncertainty creates decision-making problems but also gives rise to institutions of economic importance, like insurance arrangements. Health insurance arrangements are risk pooling arrangements that can overcome the uncertainty of immeasurable outcomes in incomparable situations, by pooling measurable risks, based on probability distributions. The next section will link uncertainty and risk pooling to problems of selection processes in unregulated health insurance markets.

2.3.2 Problems in unregulated health insurance: adverse selection and risk selection

Problems of unregulated health insurance markets can be distinguished according to their impact on efficiency and equity. Problems discussed in the literature, like adverse selection, moral hazard and supplier-induced demand, are predominantly related to increasing inefficiency due to asymmetry of information. Adverse selection means that insurers experience a disproportionate number of enrolees with predisposed high risk (Nyman 2003). Moral hazard is defined as any increase in health care consumption simply because one has become insured (Nyman 2003). Supplier or physician-induced demand means that physicians can prescribe services when they are not needed, in order to increase their incomes (Nyman 2003). Although they are typically discussed in the context of efficiency, these problems ultimately also have an impact on equity. Another problem of unregulated health insurance markets, the problem of risk selection, creates inequity. Risk selection is defined as any biased selection where certain insurers disproportionately enrol healthy, low risks, while other insurers are left with disproportionately sick, high risks (Altman et al. 1998 and Newhouse 1996).

[18] For example, Weng (1999) uses risk pooling in the context of demand uncertainty in regard to product distribution and Ramaswamy et al. (1998) uses risk pooling in the context of property ownership.

It was pointed out that early mutual health insurance arrangements entail risk pooling with cross-subsidisation. Risk pooling with cross-subsidisation occurs if each insurer accepts predictable losses on the contracts of the high-risk individuals and compensates these losses through predictable profits on the contracts of the low-risk individuals. Thus high-risk individuals would receive a subsidy from the low-risk individuals in the risk pool that increases their access to health insurance coverage. However, risk pooling realities in competitive health insurance markets impose strong incentives for insurance seekers on the demand side and for insurers on the supply side that change risk pools.

These behavioural incentives are explainable within the framework of the economic model of human behaviour. First, if the insurer cannot distinguish different risks (and charges an average or community-rated risk premia to all customers) high-risk insurance seekers[19] will have a stronger economic incentive to select themselves into insurance coverage. This adverse selection might result in unsustainable risk pooling and, in the extreme, the absence of health insurance markets. Second, if the insurer can distinguish between different risks (and charges individual risk-rated risk premia to his customers), is has an incentive to select and segment risks, and by doing so minimises the pooling of heterogeneous risks. This risk selection might result in narrow risk pooling and diminished chances for high-risk individuals to obtain or afford health insurance coverage (Holahan et al. 2003).

Further detail on the focus of this thesis, adverse selection and risk selection and their impact on efficiency and equity in unregulated health insurance markets, is presented below.

A. Adverse selection

Adverse selection in health insurance is often discussed with regard to different aspects of the information problem, mostly referring directly to asymmetry of information (e.g. Browne and Doerpinghaus 1993, Browne and Doerpinghaus 1994). The origin of this problem does not lie in incalculable uncertainty for incidents of ill-health[20], because the occurrence of illness is a stochastic or unavoidable risk that insurers can calculate. However, insurance covers not only stochastic, unavoidable risks but also deterministic or avoidable risks. The later are, for example, related to individuals' medical histories and

[19] High-risk individuals have a calculated risk (for claims) that is higher than the premium price would cover.
[20] Uncertainty is less a problem for incidents of illness than, for example, the efficacy of treatment in the case of an illness (World Bank 1993).

risk behaviours that influence their health status.[21] Knowledge about these later risks, in the context of asymmetry of information, is discussed as an information advantage of these market participants that creates inefficiency in insurance markets.

Arrow (1963) first introduced the concept of asymmetry of information to the specifics of the health sector and medical care market in his article *Uncertainty and the welfare economics of medical care.* In addition to uncertainty, *information* and *knowledge* about ill-health have to be considered when dealing with the commodities of health and health care. Arrow pointed out that: "When there is uncertainty, information or knowledge becomes a commodity." (1963:946). However, in health insurance this information or knowledge is often non-tradable and thus hard to assess or control by another party.[22] Uncertainty, asymmetry of information and the non-tradability of this information often lead to incentives incompatible with efficient market outcomes. The economic model of human behaviour can explain these incentives. For example, an individual's marginal utility can increase if the information advantage about ones' health risk is used. In health insurance markets, specific incentives for (individual) utility maximising behaviour lead to adverse selection, moral hazard[23], and supplier-induced demand[24].

Adverse selection appeared in its original definition in Akerlof (1970) as a seller's information advantage over a buyer. In the case of a (health) insurance market the opposite is assumed, that the insurance seeker has an information advantage over the insurance supplier. Adverse selection arises when insurance seekers do not act under uncertainty, but have a pretty accurate risk perception based on pre-existing medical

[21] Thus while individual protection against the risk of ill-health is limited to appropriate risk behaviour, insurance against the costs of potential health care costs is possible.

[22] In health insurance contracting information about a person's health risk, particularly in regard to that person's medical history and risk behaviour, is the exclusive and non-tradable knowledge of the (to be) insured person.

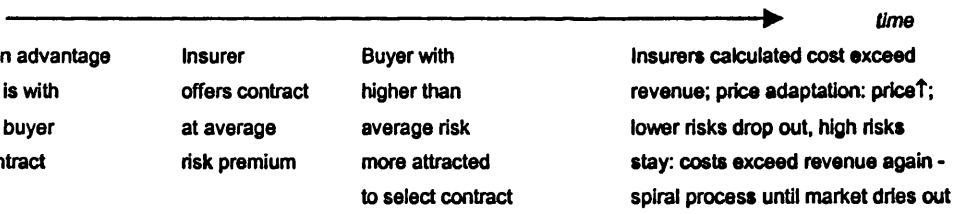
[23] Moral hazard may occur if insured use their information advantage about their actual health status against the insurer's interests and demand unnecessary or luxury health services they would not use if uninsured. Also, individuals may change their risk behaviour, i.e. insured individuals might behave in a riskier fashion simply because they are insured. Both types of individual utility maximising behaviour lead to higher levels of consumption. Insured people demand more insurance services than they would in a competitive equilibrium (under perfect information). Thus the supplying insurance company has higher costs than in a competitive equilibrium without moral hazard. With moral hazard a competitive market equilibrium (and welfare optimum) can be missed (Rothschild and Stiglitz 1976).

[24] Supplier-induced demand appears in health care provision, where information about whether a patient's health is affected takes the form of detailed and specific knowledge provided by physicians. The information advantage of the health care providers over the insurer provides an economic incentive to practise supplier-induced demand and charge more services to the insurer. Thus physicians can prescribe services when they are not needed, in order to increase their incomes (Nyman 2003). Physicians may also prescribe services when not needed or more services than are needed because they lack information or understanding about the needed care. However, both cases lead to higher health care costs for the health insurer. The effects of this supplier-induced demand are similar to the ones of moral hazard.

conditions. Thus these individuals have an information advantage over the insurer and can calculate their individual risk to a certain extent. If their calculated risk is higher than the premium price, they have a strong economic incentive to join insurance coverage. As a result of adverse selection insurers will experience a disproportionate number of enrolees with predisposed high risk (Nyman 2003).

Adverse selection is typically explained along the dimension of time, because it spans over several time-periods (see in particular: Macho-Stadler and Perez-Castrillo 1997 but also general textbooks: Nyman 2003). Asymmetry in information generally leads to adverse selection in health insurance markets because participants have a different ability to acquire information *before* a contract is signed (Cremer and Kahlil 1992, Macho-Stadler and Perez-Castrillo 1997). Thus adverse selection arises in the time-period before participants enter a contract. At this point in time the buyer has an information advantage, i.e. *ex ante* more knowledge about his/her risk than the seller - the insurance company - and can use this advantage to select the most favourable contract. If, for example, insurance seekers have an information advantage over the insurer in regard to (pre)existing health conditions or existing risk behaviour, they can impact an insurer's risk pool considerably. Figure 2-1 arranges the process of adverse selection along a time scale towards a market outcome that is inefficient and inequitable.

Figure 2-1: Adverse selection process in time dimension



The first article that analytically investigated the problem of adverse selection in the insurance market is that by Rothschild and Stiglitz (1976). By analyzing a simple model of a competitive insurance market they show that a competitive equilibrium may not exist or if it exists then only under restrictive conditions. Their model describes a market consisting of two kinds of costumers, low-risk and high-risk individuals. This market can only have either a pooling equilibrium (in which both groups buy the same contract) or a separating equilibrium (in which the two groups buy different contracts). Rothschild and

Stiglitz (1976) show that a pooling equilibrium cannot exist, because with one contract that is preferred by both risk types firms cannot break even and given a profitable contract bought by all risk types, low risks would actually prefer a different contract than the offered one. Their conclusion is that: "If there is an equilibrium, each type must purchase a separate contract." (Rothschild and Stiglitz 1976:635). However, depending on the composition of the market (i.e. if the size of the high risk group is small as compared to the low risk group) there might be a single profitable contract preferred by both types over their separated choice, upsetting the potential separating equilibrium. In that case the competitive insurance market has no equilibrium.

Since Rothschild and Stiglitz the information problem and adverse selection are frequently discussed within the context of economic efficiency (e.g. Arnott et al. 1994 and Lee 1991) and in regard to the impact on market equilibrium (e.g. Wilson 1977, Wilson 1979, Allard, et al. 1998).[25] In Cuyler's and Newhouse's (eds.) (2000) recent *Handbook of health economics,* chapter 11, Cutler and Zeckhauser (2000) discuss very detailed efficiency and equilibria in health insurance markets with adverse selection. Several sources are concerned with selection effects on market equilibrium and welfare in health insurance markets (e.g. Dahlby 1991, Belli 2001, Hansen and Keiding 2002, Kifman 2002, Danzon 2002 etc.).

Adverse selection processes create restrictions on supply and demand and the result is unsustainable risk pooling (also referred to as the "death spiral" for generous plans). If some suppliers are unable to function economically in the health insurance market they will disappear. Cutler and Zeckhauser (2000:608) point out that: "The equilibrium with adverse selection may be inefficient; it may not even exist." Thus theory predicts that adverse selection increases inefficiency in health insurance markets and, in the extreme, leads to the absence of a market.[26] However, with a variety of measures it could be attempted to work ones way back towards efficiency, potentially leading to a separating equilibrium. It should be noted that while these measures might help to induce a more efficient separating equilibrium they are not always equitable. Some of these measures are discussed below following the discussion of equity effects of adverse selection.

Adverse selection also raises equity problems:

[25] Various theoretical papers analyse market equilibrium and welfare-optimum conditions in the presence of adverse selection (e.g. Grossman 1979, Wilson 1980 and Stewart 1994).

[26] Despite the fact that there might be individuals who would like to be insured the supply of health insurance does not meet this demand. Thus the market outcome is inefficient, because it fails to successfully coordinate supply and demand.

1. Uninsured high-risk and low-risk individuals who are willing to buy insurance cannot afford the high premiums. There is income-related inequity in access to health care. Public sector arrangements of health care financing and provision might fill this gap in order to provide health care for the poor and sick (if this inequity is perceived as a societal problem). People with low income and low risk are better off in an insurance system that risk-adjusts premia. Low-income high risks would be better off in a setting where risks are pooled within a mandatory, community-rated insurance system.
2. Uninsured average and low risks, who could afford insurance and who would like to be insured choose not to, because insurance is economically unattractive compared to their risk. However, in the event of unpredicted and cost intensive ill-health, they will be unable to pay for their medical costs and become a burden for the public sector, if available.[27] This is an equity concern, since people who can afford insurance should take out insurance privately and not rely on resources that were allocated to the poor and needy. Average and low risks would be better off in a setting with risk-related premia.

Predicted insurance status outcomes under adverse selection for people with different incomes and health risks are summarised in table 2-2.

Table 2-2: Insurance patterns in a health insurance market with adverse selection

People with	Low income	High income
Low risk	Drop out because of unaffordability and economically unattractive price – add to burden of public sector if sick Equity problem: income related inequality in health care access Better off with risk-related premia	Drop out because of economically unattractive price – may become burden of public sector in catastrophic, cost intensive ill-health event Equity problem: use resources designated for the poor and needy Better off with risk-related premia
High risk	Drop out because of unaffordability – become burden to public sector Equity problem: income related inequality in health care access Better off in mandatory, community-rated insurance system	Stay in the market until it disappears (then follow high income low risks) Better off with risk-related premia or in mandatory insurance system.

There are four possible ways to achieve an efficient outcome, which separates the insured (high risks into generous plans and low risks into less generous plans): (1) charging risk related premia, (2) restricting consumer choice, (3) cross-subsidisation, and (4) risk selection in form of plan manipulation (see also Cutler and Zeckhauser 2000).

First, adverse selection can be reduced with risk-rated premiums. This is usually made possible by lifting restrictions (i.e. regulations) on how premiums are set and by allowing insurers to use information relevant for individual risk-assessment. However, this sacrifices risk spreading and forces chronically ill and high-risk people to pay higher premiums they might be unable to afford. Thus although this measure might lead to more efficiency this measure in particular sacrifices equitable access to health insurance and ultimately equitable access to health care.

Second, reducing adverse selection with restrictions in consumer choice by assigning people to different plans is possible. Insurers might offer only one insurance plan or it might be mandated that everyone within a given group of people can only buy a given plan. One example here is employment-based, mandatory insurance coverage. However, all these options reduce consumer choice and flexibility when different people prefer different plans. Additionally, restricting the timing of consumer choice by having waiting periods before someone can buy insurance or by excluding pre-existing conditions can also reduce adverse selection. However, in employment-based insurance systems this can cause people with chronic illness or high-risk dependants to be afraid to switch employers.

Third, with cross-subsidising generous plans by the less generous plans could lead to better separation of the insured. Less generous plans would be taxed with an additional per capita amount which is then used to offset the premium of the generous plan. Both insured types, low and high risks are better off with this subsidy (Cutler and Zeckhauser 2000).

Finally, a separating equilibrium can be induced by risk selection in the form of plan manipulation or plan benefit distortion. For example, the difference between two plans can be increased by converting the less generous plan into an even stingier basic plan, so that high risks more likely choose the generous plan over pooling with the low risks. However, this type of plan manipulation has a drawback on efficiency, because low risks, now preferring to pool in the basic plan, would have had higher net benefits in the previous less generous plan (Cutler and Zeckhauser 2000). Plan manipulation in form of benefit distortion is widely used in practice. Most likely there will be plans that do not offer services with predictably high costs, like oncology services. It is also possible that so designed low-risk plans include discretionary services (and sometimes luxury services,

[27] However, in many low and middle-income countries people may not fall back on the public sector for health services in the absence of private insurance. Here the reason for purchasing private insurance is often

like paramedical services) that attract the otherwise healthy low risks. Risk selection of this type and its equity effects will be discussed in more detail in the following section on risk selection.

Since unregulated market equilibria do not always lead to Pareto-optimal outcomes, one strand of literature is extensively dedicated to policy analysis and the debate on *how* to regulate markets in order to prevent selection processes or *how* to design reforms in health care/insurance markets in order to overcome undesirable adverse selection (Enthoven and Singer 1995, Blumberg and Nichlos 1996). The landmark paper by Neudeck and Podzeck (1996), for example, analyzes the effects of different policy options that correct spontaneous health insurance market dynamics.[28] They consider four cases of public provision of insurance or subsidies[29] and three cases of regulation of the private insurance market.[30]

However, there is some literature arguing strongly against regulation to improve welfare in cases of adverse selection[31] and several sources make government regulation itself responsible for market failure, the rise of adverse selection and welfare losses (e.g. Pauly 1986 on tax subsidies for health insurance). Altogether, this literature concentrates on either discussing potential policies that aim towards alternative forms of health insurance and health insurance regulation (e.g. for Latin America: Karplus and Betranou 1999, for the European Union: Mossialos and Thomson 2002) or analysing political processes of health care reforms to overcome shortcomings of health insurance market dynamics (e.g. for Southern Europe: Guillen 2002).[32] In any case, these discussions on policies for alternative health insurance systems are typically related to specific country-settings.

B. Risk selection

Risk selection in health insurance is discussed less frequently in theoretical terms as compared to adverse selection.[33] This literature is mostly connected to practical, policy-

to guarantee access to private health services in the absence of public services.

[28] They adopt Grossman's hypothesis (1979) about high-risk individuals' dissembling behavior, according to which insurers eventually turn down non-profitable contracts.

[29] These are: (1) full public insurance, (2) partial compulsory public insurance, without or with the possibility of acquiring supplementary insurance from the private sector (topping up), (3) full public insurance with the possibility of opting out and (4) risk-adjusted premium subsidies.

[30] These three cases are: (1) standard contract with full-coverage, (2) minimum insurance and (3) premium rate restrictions.

[31] Lacker (1994) questions for instance the welfare-enhancing role of government intervention derived from Wilsons' theoretical models.

[32] One widely discussed case is the development of American health benefit programs.

[33] An exception is, for example, Feldman and Dowd (2000).

oriented debates, such as the discussions of the challenges of cream-skimming in competitive health insurance markets in Pauly (1984) and van de Ven and van Vliet (1992).

Cream-skimming (also called cherry-picking) happens when unusually low risk (and low-cost) people select an insurance plan. Insurers might know more about consumers' expected costs than the consumers themselves and uses marketing or plan design to enrol a healthier-than-usual population. For example, a plan that offers excellent obstetric care but poor oncology care will probably attract a healthier population than one that offers the opposite. A common criticism of low-cost health plans is that they keep their costs low by enrolling healthier people (or encouraging unhealthy people to leave the plan) rather than by treating their enrolees more efficiently (Newhouse et al. 1999). Usually this type of plan manipulation or plan benefit distortion is initiated to combat adverse selection and promote a more efficient separating equilibrium. This also includes price differentiation between generous, high-risk plans and less generous low-risk plans. The designed low-risk plans offer not only services for predominantly healthy people, i.e. low risks, they also offer them at a much lower premium price as compared to plans designed for high risks. Offering the low-risk plan for all prospective low risks at the same lower premium might do this or, more often, by combining plan manipulation with individual risk-assessment and charging individuals risk-rated premia. This form of price differentiation is also called preferred underwriting.

Preferred underwriting creates a preferred class of risks from a group of non-rated risks using a specific set of underwriting criteria. This group of preferred risks exhibits a lower risk than the population from which it is selected and allows the insurer to offer discounted premia. Insurance brokers using underwriting criteria like medical tests, personal and family medical history, and life style considerations mostly practice underwriting.

However, risk selection is most often referred to in the context of national health care systems that use capitation systems to pay health insurers or health care providers (Barros 2003, Van Barneveld et al. 2001). Equity concerns arise here because unregulated health insurers have an economic incentive to practice risk selection, by attracting low risks and avoiding high risks (Yu et al. 2003). Risk selection is essentially any biased selection action that exploits unpriced risk heterogeneity and breaks risk pooling arrangements, with the result that some consumers may not obtain the insurance coverage they desire (Newhouse 1996).

Risk selection is practised by the insurance seller, (in contrast to adverse selection, which is practised by the insurance buyer), either to improve its risk pool and to become more profitable or in order to combat adverse selection. Thus risk selection is a process that leads to a disproportionate number of insurance enrolees and members with predisposed low risk with whom profit maximisation is possible. Risk selection is very likely to occur in contested insurance markets where several insurers compete for the most profitable low risks. First, the insurer might use signals of easily identifiable risk factors (like age, sex, race etc.) and practice preferred underwriting of potential low risks. Second, the insurer might draw on the to-be-insured's information advantage and cream-skim the market for low risks. Third, the insurer might use information and signals about identifiable risk factors (like claim and benefit patterns) to identify high risks within its existing pool and dump these by terminating their coverage. This risk selection process continues until the risk pool in that health insurance market contains only fairly good and profitable risks.

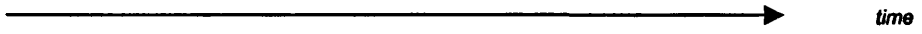
Like adverse selection, risk selection can also be very well demonstrated using the time dimension, because it spans over several time-periods (see Macho-Stadler and Perez-Castrillo 1997). Risk selection appears *before* or *after* the participants enter a contract. In the time-period *before* the contract signing the insurance seller uses signals or information about easily identifiable risk factors, like age, smoking habit, and weight of the insurance buyer to risk select.[34] This is called 'signalling' in the economics of information (e.g. Riley 1979). In this time-period the insurer also draws on the to-be-insured's information advantage by offering self-selecting plan options with risk-adjusted premiums that are particularly attractive to good risks (cream-skimming). Further risk selection is possible *after* the signing of the insurance contract. The insurer can then observe the claim and benefit patterns of its clients and terminate contracts that incur high costs, dumping these high risks. The following Figure 2-2 arranges the risk selection process events on a time scale towards a market outcome that is inequitable.

Risk selection and efficiency, especially in vertically integrated health insurance systems are connected. With increasing efficiency risk selection also rises (Newhouse 1996). An insurance plan's premium is a function of the average risk in a heterogeneous risk pool of both profitable and unprofitable (member) risks. However, integrated plans, in particular, face price competition on the cost of medical care and thus have a stronger

[34] If these risk factors are not observable, occur later in time or it is impossible to inquire about them (for instance due to legal regulation), then risk selection *before* the insurance contract is signed is limited.

incentive to produce efficiently, as well as to cream-skim (profitable) low risks and dump (unprofitable) high risks (Newhouse 1996). Because of this selection some consumers may not obtain the insurance they desire.

Figure 2-2: Risk selection process in time dimension



Information advantage/ signals about risk of insurance buyer is used by insurer *before* contract signing with risk- adjusted self-selecting insurance options (cream-skimming) or preferred underwriting	Buyer with lower than average risk more attracted to insurance contract	Signals about risk of insured *after* contract signing	Insurer excludes high risk from coverage by dumping them; proportion of good risks with of good risks in insurance pool rises - process continues until only low risks remain insured

Generally, different risk selection practices are applied in order to limit adverse selection and with it an inefficient equilibrium. Often, equilibrium may not even exist with adverse selection (Cutler and Zeckhauser 2000). While risk selection strategies might increase efficiency and promote a separating equilibrium, this comes at the cost of risk spreading (Cutler and Zeckhauser 2000). Equity in access to health insurance and ultimately access to health care will be compromised. This is discussed in the next paragraph.

Risk selection raises considerable equity concerns:

1. Uninsured high-risk individuals with low incomes, willing to buy insurance, are either excluded by the insurers as 'uninsurable' or they cannot afford the risk-adjusted insurance coverage that is offered. These people cannot afford to pay for health care out of their pockets and will experience income-related inequity in access to health care. If this inequity is perceived as a societal problem, public sector arrangement of health care financing and provision might fill this gap in order to provide health care for the poor and sick. Low-income, high risks would be better off in a setting where risks are pooled within a mandatory, community-rated insurance system.
2. Uninsured high risks who could afford insurance and who would like to be insured are also excluded as uninsurable high risks. However, in the event of cost intensive ill-health, they too may be unable to pick up their bill and become a burden on the public sector. This is an equity concern, since people who can afford insurance should take out insurance privately, and should not have to rely on resources that are allocated to the poor. High risks, (even with high income) would be better off in a setting where risks are pooled within a mandatory, community-rated insurance system.

Predicted insurance status outcomes under risk selection are summarised in table 2-3 for people with different incomes and health risks.

Table 2-3: Insurance pattern in a health insurance market with risk selection

People with	Low income	High income
Low risk	Stay in the market	Stay in the market
High risk	Drop out because of uninsurable high risk – become burden of public sector Equity problem: income related inequality in health care access Better off in mandatory, community-rated insurance system	Drop out because of uninsurable high risk –may become burden of public sector in cost intensive ill-health event Equity problem: use resources designated for the poor Better off in mandatory, community-rated insurance system

In health insurance markets risk selection can be reduced in several ways by regulation. First, non-price regulation might improve the selection-efficiency trade-off. This first type of regulation would enforce non-discrimination in risk pooling that requires cross-subsidisation, i.e. insurers have to accept high risks whose costs exceed their premia and offset those losses through profits on other insured low-risks. However, it is hard to enforce such a regulation or monitor its compliance in health insurance markets.

Risk selection also could be reduced in multiple plan choice settings by introducing (periodic) open enrolment and guaranteed renewal without medical underwriting.[35] However, insurers might alter their product to influence choice by offering specific plans services that would attract low rather than high risks, or alter incentives to gatekeeper physicians, discouraging referrals for high risks (Newhouse 1996). New plan choices are more appealing to low risks that are less likely to be established with specific physicians.

Second, there are risk adjustment strategies that consider variations in the health status of the insured in vertically integrated insurance plans in order to reduce biased selection (Newhouse 1997, Rice and Smith 2000 etc.). Risk adjustment is defined as the adjustment of premiums paid to health plans (or to insurance companies) based on a formula employing individual level diagnostic and/or demographic information (Keenan et al. 2001). Barros (2003) points out reform proposals of several national health insurance systems that have advocated variations of risk adjustment/capitation systems and face the serious obstacle of incentives to risk selection.

[35] Here anyone could enrol in any plan or remain in a plan.

Risk adjustment has been used in Belgium, The Netherlands, Israel, the United Kingdom, and Switzerland, among other countries. Thus, the literature is mostly related to specific (integrated) national health care and insurance systems (for example Israel: Shmueli 1998, Belgium: Schokkaert, and van de Voorde 2000, The Netherlands: van de Ven and van Vliet 1994, and the UK: Smith 1998). The discussions in these publications focus on adequate risk adjustment and pro-competitive regulation. However, insurers and providers can strategically play against given systems of risk-adjusted capitation payments and still risk select those consumers expected to be profitable or avoid high risks. Van de Ven et al. (1998) discuss the current difficulties in improving capitation formulas by estimating average risks and points out that the marginal improvements in the capitation formula are obtained at considerable research costs. This recent article also argues that avoidance of cream-skimming requires strong regulations and possibly some sort of mandatory high-risk pooling.[36]

2.3.3 Empirical evidence for adverse selection and risk selection in health insurance

There is a relatively limited number of studies that empirically investigate selection processes, and, of those, it is adverse selection that is predominately the centre of attention. Also, these empirical investigations are typically set in developed countries, like the US or Europe, and rarely in low- or middle-income countries. There are at least two factors that limit the number of empirical investigations into selection processes.

First, empirical studies face a lack of data. Insurance companies typically collect individual-based data for record keeping purposes. However, in low- and middle-income countries there are fewer established health insurance arrangements and data collection generally is often too costly here to be feasible. The American RAND Health Insurance Experiment is one example where data were generated from a randomised experiment in order to derive information regarding health insurance and medical care demand (and also isolate moral hazard effects) (Manning, et al. 1987, Newhouse and Archibald 1993, Manning and Marquis 1996). If insurance data is available it is not always suitable for this type of empirical investigations, because it was not collected for the purpose of identifying selection processes.

[36] See also the article of van Barneveld et al. (1996). For a good summary of the major themes of the risk adjustment literature the interested reader is referred to Newhouse (1998), Van de Ven and Ellis (2000), and van de Ven et al. (2000).

Second, empirical attempts face the difficult methodological task of identifying and isolating selection processes in practice. Measuring consequences of behaviour is a general problem in social science. Social scientists have to understand behaviour and take its motivations into account. However, they can *observe* only behaviour outcomes, not motivations. For instance, empirically observable effects of adverse selection are quite similar to moral hazard (higher health care utilisation). That makes it hard to distinguish them in data. The important distinction is that moral hazard appears if insured risks, high *and* low, have motivations to either top-up their health care utilisation or to behave more risky because they are insured. Adverse selection only concerns the sub-sample of high risks in insurance, which sought insurance motivated by predisposed high health risk and higher predictable utilisation.

The literature employs a range of different methods for the empirical identification of adverse selection and/or risk selection. It is the aim here to classify these approaches to empirical investigation into selection processes. Among the *recent* publications that conduct empirical investigations into selection processes in health insurance markets are: Sapelli and Torche (2001), Savage and Wright (2003), Soderlund and Hansl (2000) and van Vliet (2000). All four are examples of distinct methodological approaches. Table 2-4 summarises this literature, breaking the applied methods into four categories: Method (1): Multiple plan choice approach without utilisation, Method (2): Multiple plan choice approach with utilisation, Method (3): Multiple plan analysis approach, and Method (4): Risk adjustment approach. The table will summarise the studies according to the data used, the methods applied and the main results found. Finally the evidence in regard to selection will be stated.

The pages following table 2-4 will describe the four method approaches in detail. For each approach a small number of key publications will be discussed in the text. However, the majority of studies can be found summarised in table 2-4. Within the second method approach, the multiple-plan choice approach with utilisation, the American RAND Health experiment will be discussed. This section will finally discuss selected problems of these approaches.

Table 2-4: Evidence on selection in health insurance

Paper	Data	Empirical methods	Highlights of results	Selection
Method (1): Multiple plan choice approach w/out utilisation				
Propper et al. (2001)	Pseudo-panel cohort derived from repeated cross-sections of the annual UK Family Expenditure Survey 1978-1996	2 Weighted Least Square estimations for dynamics of supplementary private medical insurance purchase	Purchase of private health insurance rises with age but is less likely in older cohorts (and more likely in young). Income and quality of supplied private care are positively associated with purchase.	not explicitly concluded, however, likely
Besley (1999)	Pooled cross-section time series data from the British Social Attitude Survey, a nationally representative annual survey (1986-87, 1989, and 1990-91) of over 3,000 individuals	Descriptive statistics, probit model and a two equation model for demand of supplementary private health insurance.	Purchase of private health insurance rises with age, tailing off for those over 65. Income, house ownership and occupational attainment are positively associated with purchase, household size negatively. Among the health care quality variables only the long-term waiting list variable is statistically significant.	not explicitly concluded, however, likely
Cutler and Reber (1998)	Data of Harvard University employees after introducing a voucher-type system, with two groups of employees (2/3 that adopted system in 1995 and the remaining 1/3 that adopted it only in 1996)	Descriptive statistics and logistic regression estimates of insurance choice between plans of different generosity	Logistic model for enrolment response to the pricing reform as a function of the choice for a more or less generous plan: age, salary, being female, having faculty position and job tenure are positively associated with choosing a more generous plan. Logit models for whether people who had insurance in one year dropped coverage in the next year as functions of the plans price change and the same socio-demographic characteristic: premia significantly affect the probability of dropping coverage. Among switchers between the plans younger people move disproportionately to the less generous plan, while older people pool disproportionately in the more generous plan. Adverse selection destroyed the market for generous health insurance entirely.	adverse
Dowd and Feldman (1985)	Data from survey of twenty firms in the twin cities of Minneapolis and St. Pauls	Testing the means of several socio-demographic and health status characteristics for different health plan populations (i.e. people who select high versus low option coverage)	Individuals who chose traditional fee-for-service (high option coverage) are significantly older and more likely to self-report serious medical conditions. They are also more likely to have relatives with serious medical conditions.	adverse

cont.

cont.				
Cardon and Hendel (1996)	Data from the National Medical Expenditure Survey	Tobit model of choosing insurance versus being uninsured	Individuals who are younger, male and self-report excellent health are significantly less likely to choose to become insured.	adverse
Browne (1992)	Subsample of 2,515 group insured and 225 individually insured households from the National Medical Care Expenditure Survey 1977-1978	From a structural equation model of demand for medical insurance and expenses a group market insurance prediction equation is derived, specified with socio-economic and health status variables. The parameters of the estimated insurance prediction equation are used to predict group health insurance purchases by families with individual health insurance. Analysing the differences between predicted and actual families a demand discrepancy decomposition equation is estimated. Mean contributions of different risk groups to insurers surplus is tested for differences.	The difference between actual and predicted group health insurance purchases is greater for low risks than for high risks. This is either a sign for reduced insurance consumption of low risks and thus adverse selection or means that low risks receive more insurance through group insurance than they desire. Results of the subsidisation test shows that low risks contribute significantly more to the insurer surplus than high risks. The findings for both hypotheses together suggest that adverse selection is present in the individual health insurance market. This adverse selection leads to reduced insurance consumption by low risks and cross-subsidisation.	adverse
Bice (1975)*	A random sample of East Baltimore public housing residents	Testing the means of health status variables by different Medicaid plan enrolment, either in Medicaid fee-for-service or managed care plans	Poor reported health is positively correlated with enrolment in prepaid plans. Since the expected use of people with reported poor health is higher, there is also positive correlation with enrolment in prepaid plans.	favourable
Scitovsky et al. (1978)	Enrolment and survey data for Stanford University employees	Least-squares regression analysis of plan choice (managed care versus fee-for-service plans)	Employees who enrol in traditional fee-for-service plans are significantly older, more likely or be single or without young dependants.	adverse
Juba et al. (1980)	1976 Enrolment data and survey for Carnegie-Mellon University employees	Maximum Likelihood logit estimation for determinants of plan choice for managed care plans or traditional fee-for-service plans	Families with lower self-reported health are significantly less likely to enrol in managed care plans	adverse
McGuire (1981)	A random sample of the Yale University employees' health plan enrolment	Logistic regression of health plan choice for either managed care plans or traditional fee-for-service plans	Being female is significantly associated with joining a prepaid health plan. Age has no significant effects.	adverse
Summary of evidence for method (1)				**adverse** *cont.*

cont.

Method (2): Multiple plan choice approach with utilisation

Propper (1993)	Spring 1987 cross-sectional survey of 1,360 individuals in England and Wales	Two probit models estimating the propensity of choice for supplementary private health insurance and the propensity of captivity (=lack of choice over all possible modes under analysis) to the public health insurance	Purchase of private health insurance is most strongly associated with ability to pay and health status (besides age and risk attitude). Purchase is not significantly associated with utilisation, but there is a positive relationship between GP and inpatient utilisation and purchase. Captivity is determined by political attitude and supply-side constraints rather than health status. For example, young (<35) and old (>65) are less likely to choose private insurance. Former might be due to lack of 'need' and later due to supply-side constraints (policies not available for first time buyers in this age group and no coverage of chronic medical conditions).	adverse and favourable likely
Sapelli and Torche (2001)	Data from Chile's annual National Socioeconomic Characterisation Survey between 1990-1996	A logistic regression model of private health insurance choice including socio-economic and utilisation variables	Higher income increases the probability of being privately insured. High risks, i.e. older people and woman in the reproductive ages are less likely to be privately insured. This indicates adverse selection into the public sector or favourable risk selection into the private sector.	adverse and favourable
Sapelli and Vial (2003)	1996 data from Chile's annual National Socioeconomic Characterisation Survey	Count data models estimate utilisation equations with independent variables of number of physician visits and days of hospitalisation consumed by households in three months prior to the survey and choice of insurance; models with independent and dependant workers are estimated	For independent workers insurance plans face adverse selection and moral hazard. For dependent workers, only public insurance faces adverse selection (based on variables like age and number of dependants), private insurers not due to risk-adjusted premia. However, the results indicate that, for physician visits, moral hazard is present in both public and private insurance. There is no moral hazard in the case of hospitalisation, for either public insurance or private insurance.	adverse, favourable and moral hazard
Hansl and Soderlund (1999)	A sample of 73,000 households from the South African October Households Survey from 1995	A probit model of private health insurance choice with parameters ranging from socio-economic, health status and health care utilisation was estimated	Private health insurance coverage is significantly more likely among the formally employed, married, white, young and those who live less than 5 km away from the nearest health care facility. Higher income was significantly associated with greater probability of coverage. Significantly less likely coverage for the permanently disabled and the elderly indicate favourable selection. Adverse selection with regard to less easily identifiable medical risks is likely, with recent illness and hospital admission being associated with having taken out insurance coverage.	adverse and favourable

cont.

cont.

Hopkins and Kidd (1996)	A subsample of 16,472 households from Australia's National Health Survey of 1989-1990, a representative sample of the Australian population.	Maximum likelihood logit analysis for determinants of demand for private health insurance under Australia's universal, tax-financed public health insurance system is performed. A first specification of socio-economic and health status variables attempts to capture determinants of private health insurance purchase and a second, in which the first is nested, also includes utilisation variables.	Older people, females in the reproductive ages, families with higher income and higher the education level are more likely to purchase private health insurance. The probability of purchase is higher the more recent the last doctors visit and last hospital admittance.	adverse
Savage and Wright (2003)	A subsample of 22,913 households from the 1989–1990 National Health Survey data released by the Australian Bureau of Statistics	A probit model of insurance choice is estimated and the estimated probability of having private hospital insurance is substituted for the insurance dummy when estimating a hospital service use equation. In the regression model of length of stay the coefficient for the (in the probit model) estimated probability that an individual has private hospital insurance gives indication of moral hazard	Income and age have a significant, positive impact on the probability of having private insurance. Health status variables (captured by a detailed breakdown of chronic conditions) are found to be individually and jointly significant, providing more evidence of adverse selection. The main variable of interest of the duration of stay regressions, the estimated coefficient on the insurance variable, provide quite strong evidence for moral hazard amongst old couples and couples with dependants and weak evidence for moral hazard amongst young singles.	adverse and moral hazard
Cameron et al. (1988)	1977–1978 National Health Survey data released by the Australian Bureau of Statistics	Health care utilisation equations (negative binomial estimates) and logit models of health insurance choice (for a divided sample) using socio-economic, health status and utilisation variables	Health care utilisation is highly responsive to health status, less to age and sex, and varies little with income. Utilisation is higher for privately insured, i.e. more generous coverage, than publicly insured. Private health insurance choice is significantly increasing with age, income, education level and chronic health conditions. These results are interpreted as evidence for adverse selection and moral hazard.	adverse and moral hazard
Hurd and McGarry (1997)	A subsample of 7,327 households from the panel of the Asset and Health Dynamics Survey 1993-1994 (survey of individuals born 1923 or earlier)	Two probit models, one for elderly having supplementary private insurance and one for paying for it (with socio-economic, self-reported and other health indicator variables), two models showing the effect of private insurance, health and economic status on the probability of utilisation	Subjective health status measures have no or little effect on having supplementary private insurance, but income and wealth have are strongly associated with having private insurance. Even controlling for insurance and economic status, those in better health are more likely to purchase health insurance. Whether or not someone pays for insurance does not affect the probability of a doctor's visit or entering a hospital.	ambiguous

cont.

cont.

Wolfe and Goddeeris (1991)	Retirement Health Survey of the Social Security Administration 1969-1979 following a cohort moving into retirement	Means comparison for elderly with and without supplementary private health insurance, OLS and ln least square regression for sum of health expenditures and different types of utilisation, probit estimation of supplementary private insurance	Elderly with private health insurance had persistently greater expenditures for hospitalisation, physician visits and prescription drugs and for other measures of use showed greater utilisation, especially resource-intensive care. However, they also reported better health and more wealth. The LS regression results showed that individuals in worse health or with more wealth had higher utilisation. Past healthcare expenditures affect private insurance demand. The probit results show modest support for adverse selection as several utilisation variables predict private insurance coverage (besides being female and being wealthier). Self-reported health and disability are not significant.	adverse
Cartwright et al. (1992)	1977 National Medical Care and Expenditures Survey with data for elderly (65 or older), benefit provision data of the Health Insurance Employers' Survey linked to the National Med. Exp. Survey	Multinomial logit model for the probability of purchasing supplementary private health insurance, Two-step analysis for estimation of expected medical expenses	Purchase of supplementary private insurance is significantly more likely among females, individuals in employment, with high education and with high income. Purchase is also associated with good health, however, those in fair/poor health seem to have some difficulty in obtaining coverage, but once covered they tend to purchase comparatively more insurance. Expected medical expenses are higher for those with supplementary insurance. The results are interpreted as consistent with adverse and risk selection.	adverse and favourable
Manning et al. (1987)	RAND experiment data 1974-77 (randomly controlled trial) for the evaluation of the effects of varying generosity of health insurance coverage on health care utilisation	Means test analysis, and a four-equation model: (1) a probit model that separates user from non-users, (2) a probit model for the probability that an insured has at least one inpatient stay, given that he has other medical expenses, (3) a linear regression of the logarithm for total annual medical expenses of outpatient-users only, and (4) a linear regression of the logarithm for total annual medical expenses of any inpatient user	Use of medical services responds to co-payment levels. Medical expenses are significantly higher for plans without co-payment than the plan with co-payment. Insured individuals in a plan without co-payment are significantly more likely to use any services than insured with 95% co-payment. Use of any medical service increases with income for all plans, regardless for the co-payment level. Health status is a strong predictor for expenditure levels but not for different insurance coverage of healthy and sickly. Individuals enrolled in managed care as compared to fee-for-service plans had lower expenditure per person, but had the same rate of service use, i.e. there is less intensive care use in less generous plans.	adverse likely
Farley and Monheit (1985)	Data from 1977 National Medical Care Expenditure Survey	Ordinary least Square and 2 Stage Least Square estimation of health insurance purchase	Purchase of health insurance is significantly associated with ambulatory care expenditures.	ambiguous

cont.

cont.

Browne and Doerpinghaus (1993)	National Medical Care Expenditure Survey data 1977-1978 of the individually insured	A general linear regression model is estimated testing across six equations if lower-risk individuals purchase less complete insurance coverage than high-risk individuals	Individuals with socio-economic and health status characteristics that indicate high risk have higher utilisation (but do not purchase insurance coverage with higher cost-sharing). Adverse selection exists, where low and high risks buy a pooling policy that implicitly subsidises high risks.	adverse
Griffith and Baloff (1981)	Data of members in a start-up managed care plan located in urban St. Louis over a 5-year period	Relationship between utilisation rates and duration of membership in a managed care plan, depending on socio-economic subgrouping of the study population (sex, race, income, education and age) and for different employers and cohorts	Substantial reduction in utilisation rates with increasing duration of membership over the 5-year period in the start-up. This "start-up effect" is strongly evident across all different socio-economic subgroups, different employers and cohorts.	ambiguous
Conrad et al. (1985)	Random sample of 1980 claims and eligibility data for dental health insurance in Pennsylvania (Blue Shield)	2 and 3 Stage Least Square estimation of demand models for premia and total expenditure	Dental insurance is more likely with worse self-perceived dental health. However, risk-rating based on prior utilisation does not lead to overall premia reduction.	adverse
Ellis (1985)	Sample of employees of a large financial services firm in a single metropolitan area selecting individual health insurance coverage 1982-1983	Logit estimates of health plan choice, i.e. factors affecting managed care enrolment versus traditional health insurance enrolment over the observation period	Choice of more generous health coverage for the following year is associated with rising age and higher previous health care utilisation.	adverse
Merrill et al. (1985)	Data for enrolment and utilisation from Salt Lake City and Tallahasse	Means analysis and logistic regression analysis of health plan choice in managed care plans versus traditional fee-for-service plans, including socio-economic and health status and utilisation variables	Choice of managed care plans is associated with being younger, male and having lower utilisation (as compared to traditional insurance plan members). However, the health status variables of managed care members do not necessarily indicate low risk.	adverse or favourable likely
Wrightson et al. (1987)	Data from seven plans offering different types of managed care	Comparison of benefits for disenrolees and disenrolment rates among insured	Disenrolees are of lower risk in terms of their socio-economic group and their utilisation	adverse

cont.

cont.				
Langwell and Hadley (1989)	Medicaid consumer survey data, medical record data and secondary claims and eligibility data from Nationwide Evaluation of Medicaid Competition Demonstrations in six States (California, Florida, Minnesota, Missouri, New Jersey, and New York) 1982-1983	Person-level bivariate and multivariate utilisation and cost analyses, for example, contrasting the pre-utilisation and post-utilisation experience of a sample of managed care system enrolees in each site with pre-utilisation and post-utilisation experience of non-enrolees from comparison samples of traditional fee-for-service plans	The demonstrations - in many respects a test of the feasibility and impact of a managed care system relative to traditional, fee-for-service systems - showed significant utilisation reduction for enrolees in managed care plans on most sites. However, these apparent utilisation changes were not translated into program savings, i.e. a clear pattern of cost reductions. Studies of evidence of selection bias were undertaken for 2 sites. No evidence was detected that persons voluntarily enrolling differ significantly from those opting to remain in traditional Medicaid. Prior-use techniques and more complex econometric techniques failed to uncover bias in beneficiary plan selection.	ambiguous
Long et al. (1988)	Data of 1,553 subscribers to three Minneapolis-St. Paul managed care plans (from 27 employers) in 1984	Multivariate probit model of voluntary disenrolment, depending on premia, number of plans offered, controlled for demographic and utilisation variables	Disenrolment is a function of economic factors, disenrolment significantly rises with increase in premia and number of available plan choice. However, all demographic variables and utilisation variables in the model showed an insignificant relationship to disenrolment.	ambiguous
Ellis (1989)	Sample of employees of a large financial services firm in a single metropolitan area selecting individual health insurance coverage 1982-1983	Means analysis and 2 generalised logit models including socio-demographic and health expenditure variables by employees health care coverage choice in flexible benefit setting, 1983 introduced three new flexible benefit options with (same co-payment in 1982), but different deductibles, premia and stop-losses	Employees' choice strongly displays selection bias across plans despite similar socio-demographic characteristics. High option plan enrolees experienced significantly higher health expenditures the year prior to the change than the least generous pan option and significantly higher claims the year after the choice.	adverse
Diehr et al. (1993)	Data from interview survey among enrolees and non-enrolees (but eligibles) in Washington State's Basic Health Plan demonstration program	Multivariate analysis, including socio-economic, insurance status and health status variables of utilisation differences between enrolee and eligibles in demonstration program that provides subsidised health insurance for families earning less than 200% of the poverty level	Substantial differences between enrolees and eligibles in education, age, income, employment, race, and insurance status. In spite of these demographic and access differences, health status was remarkably similar for enrolees and eligibles, with the few significant differences favouring the enrolees. In addition, previous and subsequent use of health services was similar or lower for enrolees. The results for health status and utilisation were similar across the three counties, even though the counties and the providers were quite different.	ambiguous
Summary of evidence for method (2)				**favourable/ adverse** *cont.*

cont.

Method (3): Multiple plan analysis approach

Billi et al. (1993)	Secondary data of University of Michigan employees, dependants and retirees involved in the first open enrolment in a new university based managed care plan	Retrospective study of utilisation and demographic characteristics of those who enrolled in a managed care plan with those who did not, compare mean plan benefit payments of the managed care plan and the traditional Blue Cross-Blue Shield basic coverage plan	The group enrolled in the less generous managed care plan had a younger median and experienced lower average inpatient and outpatient payments per member in the year prior to the enrolment. These differences resulted in an overall lower payment per member for the new managed care plan group in the year prior to their enrolment.	favourable
Brown et al. (1993)	Data of Medicare insured, claims/benefits and plan enrolment in managed care plans versus traditional fee-for service plans	Comparison of average spending of switchers to managed care plans and those who remained in traditional fee-for service plans using utilisation data from two years prior to managed care enrolment	Switchers to managed care plans had two years prior to their enrolment in these less generous plans about one tenth lower spending than insured in more generous fee-for-service plans.	adverse
Altman et al. (1998)	Data on plan enrolment and utilisation for insured in Massachusetts Group Insurance Commission 1993-1996	Analyse factors accounting for differences in plan premia of managed care and traditional fee-for-service plan options, comparing average plan benefits and age- and sex-adjusted costs	Insured who switch from traditional fee-for-service to less generous managed care plans have on average one third less benefits than those retaining fee-for-service coverage. Insured who switch into the other direction used on average half more services than those remaining in managed care plans. Adverse selection results primarily form switchers not new enrolees. Adverse retention occurs, where remaining high risks magnify cost differentials between plans.	adverse and favourable
Cutler (1994)	Data from the Health Insurance Association of America, surveying yearly 3,000 firms whether coverage is provided at what premia level	Examine empirical evidence on insurance premia variability by estimating regressions for logarithm of benefit costs, depending on previous benefits, and plans' age, racial and gender composition, simultaneous equation model	Plans with more generous benefits and with a higher share of retirees and older worker have higher premia. However, premia variability is not explained by demographic differences or benefit generosity alone. There is a lack of intertemporal pooling, which might occur due to adverse selection, demand inadequacy or public subsidies to uninsured.	adverse
Buchmueller and Feldstein (1997)	Data from employees of the University of California 1993-1994 after it changed policy to limit its contribution to least expensive option	Descriptive analysis and probit models of probability for an employee to switch depending on premia level, controlling for some socio-demographic, and plan characteristics	Strong price effects were found. An increase in premia of ten USD led to five-fold increase in switchers. Increasing age and family size reduce the probability of switching to cheaper, less generous plan options.	adverse

cont.

cont.				
Price et al. (1983)	Data from the American Federal Employees Health Benefits Program, following a large premia increases in 1982-1983, and benefit reductions	Analysis of insured plan choice and comparison of premia level for plans of different generosity	The most generous plan charges premiums that are double that of the less generous plan (whereas the actuarial difference was less than ten percent). There are strong economic incentives for enrolees to leave certain high-cost plans, where the disparity between expected benefits and premium is high. The survival of some plans is questionable.	adverse and favourable
Jackson-Beeck and Kleinman (1983)	Data for employer groups health insurance coverage in Minneapolis-St. Paul (Blue Cross and Blue Shield) 1978-1981	Secondary claims analysis, comparing average health care use and costs of those who participated in first open enrolment in managed care plans with those who remained in traditional fee-for-service plans	During initial open enrolments, proportionately younger and lower users of health care enrolled in less generous managed care plans. For example, enrolees in managed care plans had on average half of the inpatient utilisation the year prior to their managed care enrolment compared with those retaining traditional coverage.	adverse
Price and Mays (1985)	Data from the American Federal Health Benefits program	Comparison of costs and premia across plan choices	Over a three-year period the most generous plans (e.g. the Blue Cross high-option) underwent a premium spiral and experience a reduction in enrolment by half.	adverse
Welch (1989)	Data from Towers, Perrin, Forster, and Crosby Inc. study of the American Federal Employees Health Benefits program	Comparison of premia of plans of different generosity (high and low option plans of Blue Cross) for government workers	The premia for the most generous plan option is about 80% higher than the least generous plan.	adverse and favourable
Cutler and Zeckhauser (1998)	Data from the Harvard University and the Massachusetts Group Insurance Commission	Analysis of premia and enrolment in plans of different generosity after changes in employers' health insurance subsidy policy	The more generous plans have higher benefits and have on average an older, high-risk membership. A premium-spiral at Harvard University led to the disappearance of the high-option plan within three years after introducing new policy. Adverse selection in the high-option plan at the Insurance Commission was contained with subsidies and tight plan management.	adverse
Soderlund and Hansl (2000)	Annual data from South Africa's 180 registered medical schemes, between 1985-1995	Descriptive statistics and three OLS models explore trends in plan enrolment and costs of private health insurance after risk-rating was legalised	A form of premium spiral is suggested, with worsening risk profiles being associated with increase in premia, loss of low-risk members and worsening of the financial position of plans. The average premia (fixed effects) model suggests risk selection, since increases in discretionary benefits, a marker of low risks, is significantly associated with lower premia.	adverse and favourable
Summary of evidence for method (3)				**favourable/ adverse** *cont.*

cont.				
Method (4): Risk adjustment approach				
Eggers (1980)	Data from a managed care Medicare risk contract with the Group Health Cooperative of Puget Sound in the State of Washington 1974-1976	Descriptive comparison of utilisation and of reimbursement between enrolees and non-enrolees in the managed care program before the open-enrolment started	Beneficiaries who joined managed care during open enrolment had half the rate of inpatient utilisation and reimbursement of non-enrolees. The open-enrolment beneficiaries' utilisation of medical services indicates a selection process. The incentive reimbursements provided to managed care plans were tied to the AAPCC, which adjusted for age, sex, residence, welfare and institutional status. However, it did not account for other factors that might affect differences in reimbursement rates, like health status.	favourable
Eggers and Prihoda (1982)	Data from three Medicare demonstration sites, including the Greater Marshfield Community Health Plan - a managed care plan	Descriptive comparison of utilisation and reimbursement patterns between enrolees and non-enrolees in the managed care program for the four years prior to the enrolment in the managed care plan	Medicare reimbursements for the four years prior to the point of enrolment were 14% lower for managed care joiners than for non-managed care plan enrolees. The Marshfield managed care plan received an AAPCC adjusted payment from Medicare, which adjusted for age, sex, race, institutional status, and disability status. The adjustment method for selection bias used has to consider the prior determined form of selection.	favourable
Dowd et al. (1995)	Data from the Twin Cities Medicare market, including five managed care plans with risk contracts, two random samples of Medicare beneficiaries (one managed care plan enrolees, one not), 1988-1989	Estimation of a model that corrects for selectivity bias (two-sector selection model), jointly estimating choice and expenditure equations (which yield selectivity-corrected tobit AAPCC expenditure equation coefficients), rather than adding more variables to the AAPCC formula and measure changes in explained variance it is attempted to measure the effect of omitted variables	The selectivity-corrected tobit AAPCC expenditure equation coefficients show that traditional (non-managed care) experience unfavourable selection. Selection on the basis of observed variables that are included in the AAPCC formula does not lead to over- or underpayments to managed care plans. However, the test for the presence of correlation between variables omitted from the choice and expenditure equations is negative and significant, meaning that these variables contribute negatively to traditional plan expenditures and indicate favourable selection.	ambiguous
Riley et al. (1996)	1994 data from the Medicare Current Beneficiary Survey, a longitudinal, a multipurpose survey of a representative sample of the Medicare population	Bivariate comparison and model estimation to predict average cost (ratio), including the health status of 863 managed care plan enrolees and 4,576 non-enrolees, controlling for demographics and area of residence	Managed care enrolees were less likely to report fair or poor health. Average predicted costs based on various health-status measures were substantially lower for managed care enrolees than for respondents in fee-for-service plans. The Medicare risk adjustment formula (AAPCC) for managed care plans does not adequately adjust for better health and consequent lower expected costs. The addition of health status measures would improve payment accuracy and significantly reduce average managed care plan payments.	favourable

cont.

cont.

Newhouse et al. (1989)	Data from the RAND Health Insurance Experiment, a sample of 7,690 person-years of those participants who completed the study	Estimation and prediction of annual expenditures for medical services with approximated variables used in the current AAPCC formula and an added set of health status and prior utilisation variables	The variance in total expenditures explained by AAPCC variables was 11 percent. Health status variables improved the explained variance by another 10-20 percent. The measures of prior utilisation substantially improved the explained variance of total expenditures. With all prior utilisation variables included 62 percent of the maximum possible variance was explained. However, the remaining one-third of the stable variation in expenditures was not picked up by the additional health status and prior utilisation variables, still leaving room for selection.	adverse or favourable
Ellis et al. (1995)	Data for a five percent Medicare sample, 1992-1992	Regression models for total 1991 Medicare program expenditures, using demographic (AAPCC), and different diagnostic and procedural variables in order to improve adjusting capitation payments to managed care plans for aged and disabled enrolees	All models with that extended the current AAPCC formula with diagnostic information predicted medical costs substantially better than the current AAPCC formula. The model that accounted for multiple medical conditions of a person achieved greatest explanatory power than models that only considered diagnostic cost groups. Prospective models predict average costs of individuals with chronic conditions as well as concurrent models.	adverse or favourable
Gruenberg et al. (1995)	Data from the 1991 Medicare Current Beneficiary Survey linked with Medicare claims, expenditures and utilisation rates	Multiple regression models predicting 1992 Medicare costs are compared, the demographic (AAPCC) model, and models added with health status variables, disability measures, a model containing a complete set of variables and a prior-use model	The comparison of the alternative risk models shows that the comprehensive and prior-use models perform and predict best. Additional variables containing direct and indirect health status measures all contribute, but to a significant different degree to forecasting health costs. All are suggested to incorporate into future AAPCC formulas. The current AAPCC formula performed worst and leaves much room for selection bias.	adverse or favourable
Weiner et al. (1995)	Data from a random sample of 624,000 Medicare beneficiaries	Comparing multiple regression models predicting total annual Medicare expenditures adding to the AAPCC formula ambulatory and inpatient diagnoses	The models that extend the current AAPCC formula with ambulatory and inpatient diagnoses predict expenditures far better. These risk-adjusted capitation payments could encourage health plans to compete on the basis of efficiency and quality and not risk selection.	adverse or favourable
Kronick et al. (1995)	Data from five State Medicaid programs: Colorado, Michigan, Missouri, New York and Ohio, 1991-1993	Regression models for expenditure predictions, using claim and eligibility data for Medicaid, including a proposed system of diagnostic categories in order to improve adjusting capitation payments to health plans that enrol people with disability	The comparison of the ratio of predicted to actual expenditures for a pure demographic model, a prospective disability payment system, a concurrent disability payment system, and a model using prior-year expenditures showed that the prospective disability payment system model predicts best. Using diagnostic groups demonstrates greater predictability of costs among people with disabilities, but leaves still room for selection.	adverse or favourable

cont.

cont.

Mellor et al. (1997)	Data of retrospectively recorded treatment from 73 randomly selected general dental practitioners in three areas in England between, 1987-1988 and 1992-1993, a random selection of their regularly attending 6-12- and 14-15-year-old patients	Mean analysis/comparison of treatment patterns of general dental practitioners in the UK working under fee-for-service in 1987-1988 with those working under capitation in 1992-1993	Dentists working under capitation in 1992-1993 were carrying out fewer examinations, fillings and extractions and were taking fewer radiographs for their regularly attending child and adolescent patients than dentists working under fee-for-service in 1987-1988. However, while treatment under capitation was lower patients received marginally more preventive care and advice. This might be interpreted as evidence for biased selection of new patients, deferment of treatment or underprovision for existing high-risk patients.	unfavourable likely
Rice et al. (2000)	Data from 8,500 general practices in England on prescription expenditures 1997-1998, 1991census data were attributed to practice lists on the basis of the place of residence of the practice population	Regression analysis for variation in age, sex, and resident originated prescribing units adjusted net ingredient cost modelled for impact of health & social needs, developing a weighted capitation formula for prescription expenditures by relating prescription costs to demographic, morbidity, and mortality composition of practice lists	A needs gradient was developed, based on permanent sickness, percentage of dependants in no career households, percentage of students, and percentage of births on practice lists. These variables (incl. supply characteristics) explained 41% of variation in prescribing cost adjusted capita across practices. This model performed better than the previous capitation formula based on previous expenditures and demographic characteristics. It formed the basis for more equitable prescription budget allocations for 1999-2000.	not applicable
van Vliet and van de Ven (1992)	Panel sample of 35,000 individuals from the largest Dutch private health insurer with expenditure and insurance coverage data for 1976-1980, additional survey data on health status indicators	Comparing performance of six varying capitation formulas for predicting individual health expenditure (i.e. explained variance), besides the basic age, sex and regional risk adjustment model they apply models with income, employment, health status or different prior health care utilisation variables	The currently used basic capitation formula systematically overpays some plans and underpays others, thus stimulating risk selection. The inclusion of prior total costs as adjusters, for example, tripled the explained variance for medical expenditure. However, because even with the inclusion of prior utilisation variables there is considerable room for risk selection it is suggested that formulas should be expanded with indicators of health status and previous diagnostic information.	favourable
van Vliet (1992)	Data of continuously enrolled in a Dutch private health insurer, 1976-1980, survey data on health status indicators	Estimating upper bound of proportion of variance in annual individual health care expenditure that is predictable, comparing predictive performance of four error component models for different capitation formulas	Poor predictive performance for the currently used basic Dutch capitation formula might induce risk selection and unfair payment of insurance plans. Incorporating prior utilisation and diagnostic health status variables into the formula could mitigate this. The best performing model (incl. these variables) predicts 20% of the variance in indiv. health care expenditures.	favourable
van Vliet (1993)	Dutch panel data for 200,000 individuals	Simulation of various alternative capitation models based on, among others, diagnostic information from previous hospitalisations	Results suggest that the problems of both risk selection and windfall profits/losses may be mitigated substantially by extending current crude capitation formulas with previous hospitalisation information and other data on prior costs.	favourable

cont.

cont.				
van Barneveld et al. (1996)	Data of 69,000 individuals insured in a Dutch insurance fund 1992-1993, cost and utilisation data	Simulation of different proposed mandatory high-risk pooling systems for predicting insurance plans profits and losses, mean analysis	The current basic capitation formula calculated inadequate premia for high-risk enrolees, providing incentive for risk selection. A system of mandatory high-risk pooling, supplementing the formula may reduce these incentives.	favourable
Lamers and van Vliet (1996)	Panel data set of about 50,000 individuals comprising annual costs and diagnostic information for 5 successive years	Comparing risk adjusting demographic and diagnostic (capitation formula) models of in their ability to predict future health care costs	Predictive accuracy of a crude age-sex-based capitation formula improves substantially when diagnostic information from individual prior hospitalisations (the longer the period over which diagnostic information is available, the better) is added. Adding diagnostic information might improve current capitation formulas that still leave room for risk selection.	favourable
Shmueli (1999)	Pooled cross-section time-series data for 4 competing Israeli health plans over 7 years, aggregated cost data and data on their population composition	Estimation of several regression models to obtain age group specific means costs and capitation rates, comparing the current capitation rates (based on age specialist visit and hospitalisation) with more comprehensive ones	The current Israeli capitation formula (based on age, specialist visits and inpatient days) underpaid plans for the middle age groups and overpaid for the elderly. The proposed capitation rate is based on estimated age specific mean costs, also including costs for ambulatory care and medicines, which play a greater role for the middle age groups. This might reflect a more appropriate and fair capitation rate, guaranteeing financial stability of funds better since they presently differ in the age composition of their members.	favourable
Shmueli (2003)	Various data sources, including the Israeli National Expenditure on Health: 1962-1999, Use of Health Services Survey 1996-1997, Sickness funds reports	Descriptive statistics for the Israeli experience of the first 6 years of risk adjustment and risk sharing (a system consisting of four components: (i) the mean normative cost level (the mean premium); (ii) the risk-adjustment scale; (iii) the payment for 'severe diseases' (risk sharing); and (iv) lump sum subsidies	Inappropriate updating of the mean premium led to deficit accumulation in the funds and to a drop in quantity and quality of care. The total number of reported 'severe disease' cases grew by 40% from 1995 to 1998, while total population grew by 9%, indicating possible gaming of the system by the funds. Selection is not assumed, although no research looks for possible evidence. Selection might become a threat if risk adjusters are not improved, and alternative risk sharing mechanisms and updating of the mean premium are discussed.	ambiguous
Schokkaert and van de Voorde (2000)	Pooled cross-section time-series data of about 100 Belgian sickness funds for 1995, socio-demographic, health status, cost data etc.	Development and estimation of a new regression model to explain medical expenditures including medical supply and other risk adjusters	The currently used Belgian risk adjustment scheme is based on different socio-demographic groups, thus leaving room for risk selection. With a risk adjustment model, adding medical supply and health status variables, large part of the variance in aggregated medical expenditure could be explained.	favourable
Summary of evidence for method (4)				**favourable**

Method (1): Multiple plan choice approach w/out utilisation. A first small group of papers analyses multiple-plan choice in health insurance without considering utilisation variables. These studies exclusively use socio-economic variables and health status variables in their choice analyses. Several papers in this group examine the determinants of an individual or household choice to obtain private (supplementary) health insurance coverage, versus relying on public health insurance coverage. These papers often do not explicitly look for empirical evidence of selection processes, but rather attempt to assess determinants of (supplementary) health insurance. However, in regard to their results it is likely that selection takes place. They include, for example, empirical investigations of data from the UK: Propper (1989, and Propper et al. 2001) and Besley et al. (1999).[37]

Propper et al. (2001) uses a pseudo-panel cohort derived from repeated cross-sections of the annual UK Family Expenditure Survey 1978-1996 to investigate the dynamics of supplementary private medical insurance purchase. The estimated demand models contain variables of age, income, cohort effects and quality of supplied care, allowing for unobserved regional differences and the effect of past purchase. They find that purchase of private health insurance rises with age, but falls across the generations. Older cohorts are less likely to purchase than younger ones. Income is positively associated with purchase. Supply variables affect the purchase of private health insurance, even after controlling for unobserved regional effects. This paper is not focused on identifying selection, but rather on examining determinants of demand for private health insurance in the UK. Thus it does not explicitly conclude selection. The results - that purchases increase with age but decrease with cohort, i.e. older cohorts are (ceteris paribus) less likely to purchase private health insurance - are explained with differences in tastes. Supported by Burchardt and Propper (1999) it is argued that those who use private care are less supportive of the equity goals of the British public health services. However, although the effect of past purchase is significant its magnitude is comparably small (Propper 2001 et al.). It is likely that the results can be interpreted in terms of selection where low risks that can afford private health insurance opt out of the public system.

Besley (1999) used a demand model similar to like Propper et al.'s (2001) ,but focused on how the demand for private health insurance depends on the inflexibility of the public sector - specifically waiting lists. He uses data from the British Social Attitude Survey, a nationally representative annual survey (1986-87, 1989, and 1990-91) of over

[37] See also the recent discussion paper from King and Mossialos (2002) on determinants of private health insurance in England.

3,000 individuals. After providing some descriptive statistics on the privately insured a simple probit model and a structural (two-equation model) for private health insurance demand are estimated. These models contain a range of socio-demographic, socio-economic, and quality of care variables. Private insurance demand can be linked to several individual and household characteristics. The results show that the demand for private health insurance rises with age, tailing off for those over 65 (presumably when premia are higher). Income, home ownership and occupational attainment are positively associated with private health insurance purchase. Larger households are less likely to obtain private health insurance. Among the health care quality variables only the long-term waiting list variable is statistically significant. These results were robust for the simple probit and the structural model. This paper focused on the results that an increase in long-term waiting lists is linked to an increased demand in private health insurance. It did not aim to identify selection. However, the results for the socio-economic model variables support an interpretation of selection effects, i.e. low risks that have enough income to afford supplementary private health insurance opt out of the public system.

Among this first group of papers is another subgroup that examines determinants of individual or household health insurance plan choice for indications (and sometimes the extent) of selection processes (for example: Cutler and Reber 1998, and Dowd and Feldman 1985).[38] These papers typically focus on the American health insurance market. Cutler and Reber (1998) investigate adverse selection effects with data on health insurance choices by employees of Harvard University. In 1995 Harvard University changed its policy from a system of subsidising generous insurance plans to a system of paying a fixed contribution independent of plan choice. As a result the price to employees for the most generous plan increased by over 500 US Dollars. The idea behind the voucher system was that it would encourage competition. People would have an incentive to search for the most efficient plan while insurers have an incentive to limit unnecessary care (Cutler and Reber 1998). However, adverse selection might occur and destroy the market for generous plan entirely. The paper of Cutler and Reber (1998) aims to estimate the welfare costs of adverse selection and then to compare these to the benefits of increased competition.

The data could be divided into a "treatment" and a "control" group - a group of 2/3 of Harvard's 10,000 employees that adopted the voucher-type system in 1995 and a

[38] Cardon and Hendel (1996) examine determinants of insurance choice versus being uninsured using data from the American National Medical expenditure Survey.

group of the remaining 1/3 that adopted the system only in 1996. As part of its analysis the paper examined the enrolment response to the pricing reform as a function of the choice for a more or less generous plan depending on a variety of socio-demographic variables (including age, sex, employee type, job tenure, salary single or family plan etc.). Age, salary, being female, having a faculty position and job tenure are positively associated with choosing a more generous plan. Further, the paper estimated logit models for whether people who had insurance in one year dropped coverage in the next year as functions of the plan's price change and the same socio-demographic characteristics. The result was that premia significantly affect the probability of dropping coverage. Looking at the age of switchers between less and more generous plans, it was shown that people who disenrol from the more generous plan are significantly younger than the ones who remained in the more generous plan. These switchers were also significantly older than people that were already in the less generous plan. As a result younger, low risks pooled increasingly in the less generous plan and older high risks in the more generous plan. The more generous plan experienced losses, subsequently increased its premium and encountered another round of nonrandom disenrolment (of younger people). Despite the high increase in the premium price, adverse selection again led to significant losses and the more generous plan was discontinued within only three years, completing the adverse selection "death spiral".

Additional literature, which documents adverse selection in multiple choice health insurance arrangements for the American market, includes, for example, Browne (1992), Bice (1975), Scitovsky et al. (1978), Juba et al. (1980), and McGuire (1981). Browne (1992) tests the hypothesis if adverse selection is present in the market for individual health insurance by comparing the amount of insurance purchased by low-risk families in the individual market with an amount that is predicted they would have purchased had they obtained their insurance in the group market. The second purpose of Browne's paper is to determine whether the equilibrium in the insurance market is a separating or pooling equilibrium if adverse selection is present in the market. The data used on this study come form the National Medical Care Expenditure Survey between 1977 and 1978. For the study a subsample of the 14,000 participating randomly chosen households was selected, 2,515 group insured and 225 individually insured households.

The first hypothesis is tested in a four-step process. First, a measure of insurance coverage is defined and, second, an insurance prediction equation is estimated for families with group coverage. This equation is used in the third step to predict the amount of

insurance purchased by individual market consumers had they acquired insurance in the group market. Fourth, the discrepancy between the insurance actually purchased in the individual market and the predicted purchases is analysed. The difference is decomposed into various factors and if the discrepancy is greater for low risks than high risks adverse selection can be concluded. The results for the demand discrepancy decomposition equation show that the difference between actual and predicted group health insurance purchases is greater for low risks than for high risks. This is interpreted as a sign for reduced insurance consumption of low risks and thus adverse selection. Alternatively, low risks receive more insurance through group insurance than they desire.

The second hypothesis of the paper is that in insurance markets a subsidisation of high risks by low risks takes place. However, the separating equilibrium definition of Rothschild and Stiglitz (1976) precludes a subsidy of this kind.[39] Comparing the mean contributions of low and high risks to the surplus of insurers tests this hypothesis. If the contribution made by low risks exceeds that of high risks there is evidence for information asymmetry and cross-subsidisation. This might indicate that the separating equilibrium of the Rothschil-Stiglitz model without cross-subsidisation describes the market for individual health insurance not very good. The results of the subsidisation test showed that low risks subsidise the consumption of high risks in the individual health insurance market. Browne's study concludes that the findings for both hypotheses together suggest that adverse selection is present in the individual health insurance market. This adverse selection leads to reduced insurance consumption by low risks and cross-subsidisation.

The here listed papers using the first method approach nearly uniformly suggest selection evidence, either for adverse selection (7 out of 11) or for favourable risk selection (1 out of 11). The three papers from the UK: Propper (1989, and Propper et al. 2001) and Besley et al. (1999) did not explicitly look for empirical evidence of selection processes, however, their results can be interpreted as selection of the young and healthy into supplementary private health insurance.

Method (2): Multiple plan choice approach with utilisation. A second strategy for empirical investigations into adverse selection is to test for a correlation between a household's choice of an insurance contract with the occurrence or severity of health care

[39] However, the models of Wilson's (1977) and Miyazaki's (1977) allow subsidisation across risk classes. Wilson's model has different assumptions about the insurance firm calculations and his model predicts as a resulting equilibrium either a separating equilibrium as in the Rothschild-Stiglitz model or a pooling equilibrium with a subsidy from low risks to high risks. Miyazaki's model assumes that firms can make

utilisation. Applying choice analysis, these studies either exclusively use variables concerning health care utilisation or a combination of socio-economic variables and health care utilisation variables. Again there are a number of empirical papers in this group that examine the relationship between individual or household health care utilisation and the choice to obtain private (supplementary) health insurance coverage, versus relying on public health insurance coverage. These include, for example, empirical investigations of data from the UK: Propper (1993), Latin America: Sapelli and Torche (2001) and Sapelli and Vial (2003)[40], Africa: Hansl and Soderlund (1999), Australia: Hopkins and Kidd (1996), and Savage and Wright (2003)[41], and the US: Hurd and McGarry (1997)[42].

Propper (1993) analyses the demand for private health insurance in the UK where the health care market is dominated by the public provider, the National Health System. She performs an estimation of insurance choice for supplementary private health insurance considering captivity (i.e. the lack of choice over all possible modes under analysis) to the public health insurance. The data used in the paper are from a cross-sectional survey of 1,360 individuals in England and Wales collected in spring 1987. Two independent probit models are estimated, one for the private insurance purchase and a second for the propensity to captivity. The propensity to captivity and purchase are both associated with income and political outlook, but differently associated with health status, health care utilisation and worry over health, attitudes towards risk, and age. Political attitude and supply-side constraints rather than health status determine captivity. For example, young (<35) and old (>65) are less likely to choose private insurance. Former might be due to lack of 'need' and later due to supply-side constraints (policies not available for first time buyers in this age group and no coverage of chronic medical conditions). Purchase of private health insurance is most strongly associated with ability to pay and health status (besides age and risk attitude), but is not associated with utilisation. While the relationship between purchase and health care utilisation is not well defined there is a positive relationship between recent GP visits and inpatient services, which might indicate that households with poor health are more likely to purchase private insurance. This could indicate adverse selection in this market (Propper 1993).

losses on individual contracts as long as they break even for the sum of all their contracts. Under these model assumptions there is a separating equilibrium with subsidisation of high risks by low risks.

[40] See also Bertranou (1998).

[41] See also Cameron et al. (1988).

[42] See also Wolfe and Goddeeris (1991) and Cartwright et al. (1992).

Sapelli and Torche (2001) study the determinants of the choice of health insurance for Chilean dependant workers and retirees. They are mandated by law to purchase health insurance, but they can choose between public and private health insurance. Privately insured have access to a wide range of private and high-quality providers while the public insurance limits access exclusively to lower quality public providers. The paper use data from Chile's annual National Socioeconomic Characterisation Survey between 1990-1996, based on random samples of households. A logistic regression model for the choice of private health insurance is estimated including socio-economic and utilisation variables. Earnings are a key factor in the insurance choice - the study confirms that higher income increases the probability of being privately insured. Additionally there are other important factors. High risks, i.e. older people and woman in the reproductive ages are less likely to be privately insured. This might indicate adverse selection into the public sector or favourable risk selection into the private sector. However, private insurers set risk rated premia while the public insurance premia are income related. While it seems likely that there is some cream-skimming by the private insurers, high-risks with low income might rationally choose inferior public insurance because they are receive a cross-subsidy here that is not portable to the better private insurance.

Sapelli and Vial (2003) also study adverse selection and moral hazard in the Chilean health insurance market by analysing the relationship between health care service utilisation and the choice of either private or public insurance. For dependent workers, who must purchase health insurance either from the one public or several private insurance providers, the relationship between health care service utilisation and the choice of either private or public insurance is analysed. In the case of independent (self-employed) workers, where there is no mandate, they analyse the relationship between utilisation and the decision to voluntarily purchase health insurance. The paper uses 1996 data from Chile's annual National Socioeconomic Characterisation Survey. Count data models estimate utilisation equations (where the dependent variables are discrete, i.e. number of physician visits and days of hospitalisation consumed by the household in the three months prior to the survey).[43] Correcting for possible selection bias, the probability

[43] The independent variables used in the utilisation equations include income, some demographic characteristics, and a dummy on self-reported illness or accident during the previous three months. Affiliation to a health insurance plan is utilised as a proxy for the fees charged for services. The choice of health insurance plan, in turn, is a function of the individual's income, the premium charged for each insurance plan, and the demographic characteristics of the individual and his or her dependants.

of purchasing each type of insurance is estimated. The choice of health insurance is made by the household presumably prior to the three-month period. The model is used to detect over-utilisation (i.e. moral hazard) associated with the purchase of one kind of insurance.[44]

For independent workers, they find that insurance plans in Chile receive an adverse selection of the population, and their affiliates consume more than they would have if they had not purchased insurance (moral hazard). When analysing the choice of public and private insurance for dependent workers, they find that adverse selection is present only against public insurance based on variables like age and number of dependants, but not against private insurance. This can be explained with the fact that private insurance has risk-adjusted premia, while public insurance does not. However, for physician visits, moral hazard is present in both public and private insurance. There is no moral hazard in the case of hospitalisation, for either public insurance or private insurance. This result is intuitive, since the price elasticity of demand for hospitalisation is low.

Hansl and Soderlund (1999) present an empirical analysis of the determinants of private health insurance coverage in South Africa. South Africans are either covered by the general tax-financed public health system, which is of poor quality or may purchase private insurance, which guarantees access to high-quality private health care. The data used is a sample of 73,000 households from the South African October Households Survey from 1995. A probit model of private health insurance choice with parameters ranging from socio-economic, health status and health care utilisation is estimated. The results show that individuals are more likely to have health insurance coverage if they are formally employed, married, white and young. Those who lived more than 5 km away from the nearest health care facility are less likely to have coverage. Higher income is significantly associated with greater probability of coverage.

Significantly less likely coverage for the permanently disabled and the elderly indicate that risk selection with regard to "easily identifiable" risk factors such as age and disability, was operating. At the same time, there were suggestions of adverse selection with regard to less easily identifiable medical risks in operation, with recent illness and hospital admission being associated with having taken out insurance coverage. Relatively high numbers of middle-aged and elderly people able to afford an average private health

[44] Altogether six models are estimated: the utilisation of physician visits and hospital days within three population groups, holders of a private insurance policy and of public insurance in the case of dependent workers, and members of a health insurance plan (be it private insurance or public insurance) in the case of independent workers.

insurance premium, appear not to have coverage, and a large proportion of these must be forced to rely on the state in the event of serious illness. Taken together with the fact that both elderly and young adults had lower levels of coverage, these figures suggest that there are significant levels of both adverse selection and cream-skimming/dumping in the South African health insurance market. The significant unravelling of insurance risk pools in South Africa over the 1990s, with consequent loss of private health insurance coverage for high risk groups placed an increasing burden on the poor public system. While there have probably been a number of causes, deregulation of the industry in the late 1980s is likely to have been a major contributing factor. Extensive risk rating of medical scheme premiums resulted in inequities in access to insurance coverage.

Hopkins and Kidd (1996) investigate the determinants of demand for private health insurance under Australia's universal, tax-financed public health insurance system. Private health insurance gives members the choice of their own doctor in public hospitals and subsidises the cost of private hospital care. It also subsidises ancillary care, like physiotherapy and dental care. The paper uses a subsample of 16,472 households from Australia's National Health Survey of 1989-1990, a representative sample of the Australian population. A logit analysis is applied, with a first specification of socio-economic and health status variables attempting to capture determinants of private health insurance purchase and a second, in which the first is nested, including additional utilisation variables. The results show that the age groups beyond the age of 35 are more likely to purchase private health insurance. The probability of purchase is higher the more recent the last doctors visit and last hospital admittance, the higher the families income, and the higher the education level. Females in the reproductive ages are also more likely to purchase private health insurance. The results on the whole confirm that private health insurance experiences adverse selection, because enrolees are of higher risk.

Savage and Wright (2003) discuss adverse selection and moral hazard with 1989-1990 National Health Survey data released by the Australian Bureau of Statistics containing information of 22,913 households on private hospital usage. They analyse the relationship between insurance coverage choice and hospital service usage. In private hospitals individuals pay for services, while services in public hospitals are free to all, but are delivered after a waiting time. Thus the decision to purchase insurance for private hospital treatment depends on the trade-off between the price of treatment, waiting time, and the insurance premium. A theoretical model is developed which provides reduced form equations for insurance choice, hospital choice, and hospital use. A probit model of

insurance choice is estimated and the estimated probability of having private hospital insurance is substituted for the insurance dummy when estimating the hospital service use equation.[45] The regression model is estimated with the log of length of stay as the dependent variable, the covariates are reason for hospital stay, the estimated probability that an individual has private hospital insurance (obtained from the insurance choice probit analysis), and some socio-economic variables, including region of residence. The estimated coefficient on the estimated probability of insurance variable gives an indication of moral hazard.

In all probit models of private insurance choice, income and age have a significant, positive impact on the probability of having private hospital insurance. Health status variables (captured by a detailed breakdown of chronic conditions) are found to be individually and jointly significant, providing more evidence of adverse selection. The main variable of interest of the duration of stay regressions is the estimated coefficient on the insurance variable, which gives an indication of the extent of moral hazard. The results provide quite strong evidence for moral hazard amongst old couples and couples with dependants and weak evidence for moral hazard amongst young singles.

Hurd and McGarry (1997) investigate the influence of private supplementary health insurance on health care utilisation for the American elderly covered by Medicare. Medicare covers 95% of all elderly (65 or over) and pays for approximately half of their medical expenses. Most elderly (75%) buy private insurance or receive coverage from former employers' retirement packages for services and costs not covered by Medicare. They use a data subsample of 7,327 households from the Asset and Health Dynamics Survey 1993-1994. They first estimate a probit model for the probability of having private insurance. This only includes socio-economic, self-reported health and other health indicator variables. They find that income and wealth are strongly associated with having private insurance, but most subjective health status measures have little or no effect on private insurance status. A second model estimates the probability of paying for private insurance as a function of the same variables. Even controlling for insurance and economic status, those in better health are more likely to purchase health insurance. Finally, two models are estimated, showing the effect of private insurance, health and economic status

[45] Separate probit models were estimated for income units with different compositions (singles, couples and couples with dependants) to capture their heterogeneity. The single and couple income units were separated into young (head aged less than 50 years) and old

on the probability of (1) seeing a doctor and (2) staying overnight at a hospital. Whether or not someone pays for insurance does not affect the probability of a doctor's visit or entering a hospital. Thus none of the results provide evidence for adverse selection.

A subgroup among this second group examines the determinants of health service usage given that households are insured (Manning and Marquis 1996, Manning et al. 1987) and specifically the choice to become insured in plans of different levels of coverage (for example, Browne and Doerpinghaus 1993). This literature exclusively concerns the American health insurance market.

The relationship between insurance and service use is well documented in studies based on the American RAND Health Insurance Experiment (Newhouse et al. 1993). The RAND experiment was a randomly controlled trial to evaluate the effects of varying generosity of health insurance coverage between 1974-1977. A very detailed description of the study design can be found in Newhouse et al. (1981). Here only some central features of the study will be noted. Families participating in the RAND experiment were randomly assigned to health insurance plans that varied in co-payments and deductibles. These families came from four metropolitan areas and two rural areas, 70% of them participated three years and the remaining 30% five years.[46] The different levels of co-payment were 0, 25, 50 or 95%, the different levels of deductibles were 5, 19 or 15% of the family income or a fixed amount of 1,000 US Dollars. One plan with 95% -payment rate had a deductible of 150 US dollars per family member (max. 450 US Dollars). At the end of the study each family was hypothetically offered supplementary insurance coverage for reducing their deductibles by either one third, two thirds of 100% at a given (randomly generated) premium. The results of the experiment showed that greater cost-sharing for the insured decreased health expenditures. Families with the largest cost-sharing had up to 30% less than those without any cost-sharing.

Manning and Marquis (1996)[47] and Manning et al. (1987) use RAND experiment data in order to analyse issues in health insurance demand and health insurance

---

partitions and separate probit models were estimated for each, the rationale being that young and old, singles and couples are very distinct groups with different behaviour.

[46] Families headed by an individual over 62, individuals eligible for other State health programs or families within the upper 7% of the national income distribution were excluded.

[47] Manning and Marquis (1996) examine the trade-off between risk pooling and moral hazard by estimating both the demand for health insurance and the demand for health services. The paper uses for their estimation of the demand elasticity a utility function and applies it to RAND total expenditure data in the 25-95% co-payment range. The equation system[47] uses as dependent variables use of health services and insurance choice and independent variables insurance plan co-payment levels, health status, anticipated health care expenditures and socio-economic variables. The demand for health care was found to be significantly

utilisation. Their general findings are that a household's health service consumption is higher the more health insurance coverage they have, the greater their income, and, to some extent, the worse their health status.

Manning et al. (1987) analyse health services utilisation for families insured in plans of different generosity (i.e. cost-sharing) controlling for socio-economic, health status and regional variables. They use means test analysis and a more precise four-equation model. The first equation is a probit model that separates user from non-users, describing their characteristics. The second probit model estimates the probability that an insured has at least one inpatient stay, given that he has other medical expenses. The third model is a linear regression of the logarithm for total annual medical expenses of outpatient-users only. The fourth is a linear regression of the logarithm for total annual medical expenses of any inpatient user. The results show that the use of medical services responds to co-payment levels. Medical expenses are significantly higher for plans without co-payment than the plan with co-payment. Insured individuals in a plan without co-payment are significantly more likely to use any services than those insured with 95% co-payment. Use of any medical service increases with income for all plans, regardless for the co-payment level. Health status was a strong predictor for expenditure levels but not for different insurance coverage of healthy and sickly. Individuals enrolled in managed care as compared to fee-for-service plans had lower expenditure per person, but had the same rate of service use. The results of less intensive care use in less generous plans might indicate adverse selection.

The number of papers concerning the analysis of selection processes in the American health insurance market with real-world data is substantial. These are, for example, Browne and Doerpinghaus (1993), Griffith and Baloff (1981), Farley and Monheit (1985), Conrad et al. (1985), Merrill et al. (1985), Wrightson et al. (1987), Langwell and Hadley (1989), Long et al. (1988), Ellis (1985 and 1989) and Diehr et al. (1993).[48] Their data,

associated with price and income, the demand being price and income elastic. The results for the health insurance demand show constant absolute risk aversion in income. The estimated demand function is used to calculate expenditures for health care and the associated deadweight loss from moral hazard. They conclude that optimal plans with coinsurance only would have a co-payment level of around 45%. The rate at which the marginal gains from increased risk-pooling equals the marginal loss from increased moral hazard. The results for an optimal plan with a deductible or maximal out-of-pocket payment (stop –loss) are less clear, indicating possibly a very high stop-loss, which seems not plausible.

[48] See also Hill and Brown (1990). They study choices of Medicare beneficiaries to enter different plans by comparing health care utilisation data prior individual's choice of plans of different generosity levels and analyse mortality data for all plans in the year following the choice. After adjusting for some socio-economic variables, individuals, who chose less generous plans in the following year, had significantly less prior service usage. Mortality in the less generous plans is also significantly lower the year after the plan choice.

empirical methods, and main results are summarised in table 2-4. All the studies named here analyse the choice to become insured in plans of different levels of coverage including utilisation variables (of some kind).

Browne and Doerpinghaus (1993) analyse data from the National Medical Care Expenditure Survey 1977-1978 of the individually insured. The study extends empirical investigations of the individual medical insurance market and tests whether there is reduced consumption of insurance by low risks, whether a pooling or separating model better characterises the market, and whether cross-subsidisation from low to high risks occurs. A general linear regression model is estimated, testing across six equations whether low-risk individuals purchase less complete insurance coverage than high-risk individuals. The study finds that individuals with socio-economic and health status characteristics that indicate high risk have higher utilisation (but do not purchase insurance coverage with higher cost-sharing). They conclude that adverse selection exists, where low and high risks buy a pooling policy that implicitly subsidises high risks. Low risks consume less insurance than in a market without adverse selection. Thus the results support a pooling rather than separating equilibrium.

The majority of the discussed studies using the second method approach suggest selection evidence, either for adverse selection (17 out of 22) or for favourable risk selection (7 out of 22). Evidence for moral hazard was found in 3 out of the 22 studies. Six publications in the list of 22 reported ambiguous or no evidence for selection. However, studies with this outcome concerned only the American health insurance market and were often confined to demonstration settings (e.g. Langwell and Hadley 1989 and Diehr et al. 1993) or managed care start-ups (e.g. Griffith and Baloff 1981).

Method (3): Multiple plan analysis approach. A third, smaller group of studies analyses exclusively aggregated, i.e. plan-level, data of multiple plan options for signs of selection processes. These studies range from descriptive comparisons of some depicted plan characteristics (like average plan age, plan costs and premia) to statistical modelling of plan cost functions or plan risk profiles. Most of these studies relate to the American health insurance market, for example, Billi et al. (1993), Brown et al (1993), Altman et al. (1998), Cutler (1994), Buchmueller and Feldstein (1997), Price et al. (1983), Jackson-Beeck and Kleinman (1983), Price and Mays (1985), Welch (1989), and Cutler and Zeckhauser

They conclude that healthier individuals choose less generous plans and that plans profit from enrolling low risks.

(1998). While two of these studies will be discussed in more detail here in the text, the others are shortly described in table 2-4.

Billi et al. (1993) conduct a study to determine whether favourable or adverse selection occurred in a preferred provider organisation (PPO) enrolment, a form of managed health care, which can be describes as a less generous plan. They use secondary data in a retrospective study of the utilisation of health services and demographic characteristics of the population involved in the first open enrolment in a new university-based PPO. The PPO under study, sponsored by the University of Michigan Medical Center, was offered to all its 43,005 employees, dependants, and retirees. Prior to the start of the PPO almost all employees had traditional Blue Cross-Blue Shield basic coverage and major medical coverage through a separate company. They analysed insurance company payments during the one-year period prior to the enrolment to compare the utilisation patterns of those who enrolled in the PPO with those who did not. The groups enrolled in the less generous PPO had a younger median age than the non-PPO group; the sex distribution was roughly similar for the two groups. The PPO group experienced lower inpatient and outpatient payments per member in the year prior to the enrolment. These differences resulted in an overall lower payment per member for the PPO group in the year prior to their enrolment. They conclude that the less generous PPO received favourable selection during the open enrolment.

Altman et al. (1998) analyse factors accounting for differences in plan premia of managed care and traditional fee-for-service plan options, comparing average plan benefits and age- and sex-adjusted costs. They use data on plan enrolment and utilisation for people insured through the Massachusetts Group Insurance Commission between 1993-1996. The Massachusetts Group Insurance Commission provides health insurance to roughly 133,000 state and local employees and their families, making it the largest insurance purchaser in New England. The purchaser offered several health insurance plans. The most generous plan is a traditional fee-for-service plan, followed by a PPO plan with network providers and mild utilisation restrictions. The ten most stringent plans comprise HMO plans. The bulk of the insured enrolled in the one traditional plan or the ten managed care plans. Altman's et al. (1998) results show that insured who switch from traditional fee-for-service to managed care plans have, on average, one third less benefits than those retaining fee-for-service coverage. Also, those insured who switch into the other direction, i.e. from less generous managed care plans to fee-for-service plans, used, on average, half more services than those remaining in managed care plans. They

conclude that adverse selection results primarily from people switching plans and not new enrolees disproportionately moving to certain plans, as the classical adverse selection cycle suggests. Thus they talk instead about 'adverse retention', defined as the process where remaining high risks magnify cost differentials between plans.

While the previous examples are exclusively investigations into the US health insurance market, Soderlund and Hansl (2000) explore trends in plan enrolment and costs of private health insurance for South Africa's medical schemes after risk-rating was legalised. They investigate two hypotheses, first, risk distribution is unequal between health insurers and increases over time, and second, risk selection leads to less efficient health care spending. Annual data from all (180) registered South African medical schemes, including membership information, premia and benefits (by type) and other financial plan data between 1985-1995 are used. Soderlund and Hansl (2000) use descriptive statistics and estimate three OLS models. One model illustrates the association between the plans' pensioner ratio, its organisational nature and time. The two other models (one fixed effects) estimate the association of plan premia with several plan characteristics (i.e. plan risk and size, plan type etc.).

The results show adverse selection evidence, suggesting a form of premium spiral, with worsening risk profiles being associated with increases in premia, loss of low-risk members and worsening of the financial position of plans. However, results from the third fixed effects model suggest risk selection, since increases in discretionary benefits, a marker of low risks, is significantly associated with lower premia. Also, it can be shown that risk-rating (and lower costs) for low risks is accompanied by quality improvement, i.e. moral hazard related higher levels of benefits rather than premia reduction. Soderlund and Hansl (2000) verify the hypothesis that there has been a definite separation of risks between plans over the study period, although it cannot be explicitly attributed to the policy change of risk-adjusted premia. It is also concluded that risk-rating, which can be directly attributed to the policy change, is associated with increased premia costs (and thus worsened inefficiency).

The discussed papers using the multiple-plan analysis approach uniformly suggest selection evidence, either for adverse selection (10 out of 11) or for favourable risk selection (5 out of 11).

Method (4): Risk adjustment approach. A fourth type of empirical literature is described as the risk adjustment approach. Risk-adjustment can either mean to group reasonably homogeneous individuals in insurance plans or to adjust plans' premium

revenues (Newhouse 1996). Studies in this group most often discuss biased selection in the context of so-called capitation systems of national health insurance systems or other vertically integrated systems. In capitation systems health insurance plans or health care providers receive a payment from the government or private payers for each enrolee or registered patient. Many countries turned to capitation or competition among managed care plans in order to overcome the trade-off between cost and quality in health care provision. (Frank et al. 2000).[49] In the US, major public health insurance programs, like Medicare, and many private health insurance plans offer enrolees a choice of managed care plans paid by capitation. In other countries with national health care systems, like Israel, the Netherlands, and England, health policy shares similar features. Israel, for example, reformed its health care system so that residents may choose among several managed care plans which all must offer a comprehensive package of health care services set by the regulator.

However, insurance plans paid by capitation have an incentive to distort the quality of services they offer to attract profitable and to deter unprofitable enrolees (Frank et al. 2000). That means plans will most likely practice favourable or unfavourable risk selection, partly to prevent adverse selection, or, as Miller and Luft (1997:20) put it: "Under the simple capitation payments that now exist, providers and plans face strong disincentives to excel in care for the sickest and most expensive patients. Plans that develop a strong reputation for excellence in quality of care for the sickest will attract new high-cost enrolees...". This risk selection manifestation in the form of underprovision of care in order to avoid bad risks is called by Ellis (1998) 'skimping', Newhouse et al. (1997) 'stinting', while Cutler and Zeckhauser (2000) refer to it simply as 'plan manipulation'.

Thus the most common goals of studies in this group are the development and the improvement of formulas for health insurance plans' premia adjustments in order to reduce insurance plans' incentive to practice risk selection or to compensate plans appropriately that offer expensive and extensive care for the sickest. These risk adjustment formulas are based on insurance enrolee characteristics, including socio-economic, health status and health care utilisation information.

[49] The capitation or managed care strategy is based on the assumption that costs are controlled by the capitation payment while the quality of services is enforced by the market. Thus the capitation payment gives plans an incentive to reduce costs (and quality), while the opportunity to attract enrolees gives plans an incentive to increase quality (and costs). "Ideally, these countervailing incentives lead plans to make efficient choices about service quality" (Frank et al. 2000:829).

Typically the risk adjusters attempt to minimise the time invariant, subject-specific effect in an equation that seeks to explain variance in annual individual medical expenditures. There is extensive research on the contribution of various adjusters to the explained variance (for example: Newhouse et al. 1989, Newhouse and Archibald 1993, Epstein and Cumella 1988). Socio-economic adjusters such as age, race and geography generally explain little or modest variance (Newhouse 1996). Information on health status and chronic conditions explain modest variance, but they are based on self-reported measures that are hard to audit and are susceptible to gaming (Brown et al. 1993). Prior utilisation explains most variance (Ash et al. 1989, Newhouse 1996). However, risk adjustment methods are still far from perfect and many empirical studies found evidence for selection.

Chernichovsky and van de Ven (2003:1) summarise that "the practice of risk adjustment allocation and reimbursement started thirty years ago in the United States (US). Since 1972 Health Maintenance Organizations (HMOs) in the US have had the option to risk contract with Medicare, the social health insurance for the elderly. For each Medicare patient the HMO then received an adjusted average per capita cost (AAPCC)-capitation payment. This payment is based on the AAPCC, which is supposed to represent the local per capita fee-for-service costs that would have been expended, if the individual Medicare recipient had remained in the fee-for-service medical system instead of enrolling in the HMO. Until 2000 the AAPCC-capitation was adjusted for the age, gender, region, institutional status and welfare status of the individual. In the 1990s risk adjustment has been implemented in the health insurance system in at least ten other countries, in particular in Europe."

Articles of this fourth type of literature include US studies on risk adjustment for HMOs risk contracts with the federal Medicare insurance program for the elderly and other capitated delivery systems. These studies generally have two aims. First, they investigate evidence for selection bias into Medicare HMOs and commonly conclude that the adjustments embodied in the AAPCC formula - age, sex, welfare status, and institutional status - are too crude. Studies of this group include, for example, Eggers (1980), Egger and Prihoda (1982), Dowd et al. (1995), and Riley et al. (1996). Second, they discuss improvements in the AAPCC formula, for example, Newhouse et al. (1989), Ellis et al. (1995), Gruenberg et al. (1995), and Weiner et al. (1995)[50]. From this extensive body of

[50] See also Kronick et al. (1995) on a system of diagnostic categories that Medicaid programs can use for adjusting capitation payments to health plans that enrol people with disability. Frank et al. (2000) assess

literature for the American health insurance market, two articles will be discussed in more detail here. Others are shortly described in table 2-4.

Riley et al. (1996) investigate the health status of Medicare enrolees in HMOs and fee-for-service plans in order to derive evidence for selection processes. They compared the health status of 863 health maintenance organization (HMO) enrollees with that of 4,576 non-enrollees, controlling for demographics and area of residence, using 1994 data from the Medicare Current Beneficiary Survey (MCBS). HMO respondents were less likely to report fair or poor health, functional impairment, or heart disease. Average predicted costs based on various health-status measures were substantially lower for HMO respondents than for respondents in fee-for-service (FFS) arrangements. The Medicare payment formula for HMOs does not adequately adjust for the better health and consequent lower expected costs of HMO enrollees. The addition of health-status measures would improve payment accuracy and reduce average HMO payments significantly below current levels.

Newhouse et al. (1989) estimates how well proposed additional risk adjusters for the AAPCC formula based on health status and prior utilisation predict annual medical expenditures among non-elderly. They use data from the RAND Health Insurance Experiment, a sample of 7,690 person-years of those participants who completed the study. The authors estimate and predict annual expenditures for medical services with approximated variables used in the current AAPCC formula and an added set of health status and prior utilisation variables. The variance explained by AAPCC variables was only 11 percent of the variance in total expenditures. Health status variables improved the explained variance, raising the percentage of explained variance in total expenditures to 20-30 percent. The measures of prior utilisation lead to a substantial improvement in the explained variance of total expenditures. The prior-year utilisation variable alone rose the explained variance to 44 percent, and with all prior utilisation variables included 62 percent of the maximum possible variance is explained. However, the remaining one-third of the stable variation in expenditures was not picked up by these additional health status and prior utilisation variables, leaving room for selection.

adverse selection in managed care with data of health claims and enrolment files from the Michigan Medicaid program for the years 1991–1993. They combine several factors (dispersion in health costs, individuals' forecasts of their health costs, the correlation between use in different illness categories, and the risk adjustment system used for payment) in an empirically implementable index that can be used to identify services that will be most distorted by selection incentives. They find that information (individuals' knowledge and their ability to forecast their health expenses) has a dramatic effect on selection incentives and

Risk adjustment or capitation payments are also used in Belgium, The Netherlands, Israel, the United Kingdom, and Switzerland. Thus part of the risk adjustment literature is related to specific (integrated) national health care and insurance systems in these countries. Van de Ven et al. (2003) discuss risk selection in the sickness fund insurance markets of five European countries, Belgium, Germany, Israel, the Netherlands and Switzerland. They present a conceptual framework for understanding risk adjustment and compare risk adjustment systems of the five countries. The article concludes that in the case of imperfect risk adjustment—as is the case in all five countries in the year 2001—the sickness funds have financial incentives for risk selection, which may threaten solidarity, efficiency, quality of care and consumer satisfaction. It is expected that without substantial improvements in the risk adjustment formula, risk selection will increase in all five countries. Other articles focus on country-specific empirical analyses of data, for example, for The Netherlands, van Vliet and van de Ven (1992 and 1993), van Vliet (1992), van Barneveld et al. (1996 and 1998)[51], Lamers and van Vliet (1996), van Vliet (2000), van Barneveld (2001)[52]. Recent articles for the UK include Mellor et al. (1997), Gravelle et al. (2003)[53], and Rice et al. (2000). Shmueli (1999) and Shmueli et al. (2003) use data for Israel and Schokkaert and van de Vorde use Belgian data.[54]

From the equally extensive body of risk adjustment literature, mainly related to the European health insurance market, four articles, will be discussed here, while others are shortly described in table 2-4. The first article by Mellor et al. (1997) assesses the possible impact of capitation payment systems on service delivery of general dental practitioners in the UK. The second article by van Vliet and van de Ven (1992) contains an empirical analysis with Dutch panel data to show that too crude risk adjusters (on which capitation

if people know what they are commonly assumed to know (age, sex and prior spending), selection incentives would be very severe.

[51] The studies of van Barneveld et al. (1996) and van Barneveld et al. (1998) essentially have the same purpose, assessing forms of mandatory pooling as a supplement to risk adjusted capitation payments. The article of van Barneveld et al. (1996) is described in table 2-4. Van Barneveld et al. (1998) uses data of enrolees in a Dutch insurance fund during 1988–1991. Different variants of mandatory pooling, high-risk pooling with excess-of-loss and proportional pooling are compared. Each variant includes *ex post* compensations (to insurers for some members), depending to various degrees on actually incurred costs. All pooling variants reduce the incentives for risk selection which are inherent in the imperfect Dutch capitation formula, but they also reduce the incentives for efficiency. They analyse which of the three main pooling variants yields the greatest reduction of the risk selection incentives, given a measure for the incentive for efficiency (% of total costs for which an insurer is at risk). The high-risk pooling is the most effective.

[52] Van Vliet (2000) and van Barneveld et al. (2001) have a similar theme like van Barneveld et al. (1996). These two first named articles are not listed on table 2.4 but in the following text.

[53] Gravelle et al. (2003) model supply and demand influences on the use of health care with implications for deriving a needs based capitation formula.

[54] See also Beck and Zweifel (1998) who present evidence for cream skimming in Switzerland's deregulated social health insurance.

payments are based) provide room for risk selection by the insurers. The two articles of Vliet (2000) and van Barneveld (2001) assess forms of mandatory pooling as a supplement to risk adjusted capitation payments in The Netherlands.

Mellor et al. (1997) compare the patterns of treatment of general dental practitioners in the UK working under fee-for-service in 1987-1988 with those working under capitation in 1992-1993. They analyse data of retrospectively recorded treatment from 73 randomly selected general dental practitioners in three contrasting areas in England during 1987-1988 and 1992-1993. They use data from a random selection of regularly attending 6-12- and 14-15-year-old patients. The results of mean analysis show that for example, the mean numbers of examinations per year reduced in the three areas from 1.75 in 1987-1988 to 1.3 in 1992-1993. Mean numbers of visits per patient dropped from 2.7 to 2. Altogether, in 1992-1993, dentists working under capitation were carrying out fewer examinations, fillings and extractions and were taking fewer radiographs for their regularly attending child and adolescent patients than dentists working under fee-for-service in 1987-1988. However, while these patients attended less frequently for treatment they also received marginally more preventive care and advice. For example, mean percentages of children per dentist receiving oral hygiene instruction rose from 18-31% in 1987-1988 to 26-33% in 1992-1993. Dietary advice increased from 3-18% to 11-20%. The result of less visits and treatment events might be interpreted as evidence for biased selection of new patients, deferment of treatment or underprovision for existing high-risk patients. However, the increase in preventive care and advice might indicate a successful cost containment strategy.

The article by van Vliet and van de Ven (1992) first has the objective of showing that the capitation formula of the Dutch national health insurance system is too crude, leading to unfair distribution of funds over insurers and leaving room for risk selection by the insurers. Second, it is the objective to investigate more comprehensive formulas that may reduce these problems. They use a panel data sample of about 35,000 insured individuals from the largest Dutch private health insurance organisation containing expenditure and insurance coverage data for the years 1976-1980. Additional survey data on health status indicators was available and used. They compare the performance of six capitation formulas including a varying number of risk adjusters for predicting individual health expenditure. Besides the basic age, sex and regional risk adjustment model they apply models with income, employment, health status or different prior health care utilisation variables. The results are in many respects in line with findings of similar

studies employing US data. First, the currently used basic capitation formula systematically overpays some and underpays others, thus stimulating risk selection. The inclusion of prior total costs as adjusters, for example, tripled the explained variance for medical expenditure. However, because even with the inclusion of prior utilisation variables there is considerable room for risk selection it is suggested that formulas should be expanded with indicators for health status and previous diagnostic information.

Several papers suggest, in case risk-adjusted capitation payments cannot be improved in practice, that insurers' incentive to risk select can be reduced by supplementing the capitation payment with a form of risk sharing (or also mandatory high-risk and high-cost sharing) between the insurer and the regulator (for example: van Barneveld 1996, van Barneveld et al. 1998 and 2001, and van Vliet2000). Risk sharing means that insurers are retroactively reimbursed by the regulator for some of the expenditures of their members. Van Vliet (2000) and van Barneveld et al. (2001) aim to develop and assess forms of mandatory pooling as a supplement to risk adjusted capitation payments in The Netherlands. All these papers start out from the point that the currently used Dutch age-sex capitation formula leaves too much room for risk selection by sickness funds.

Van Vliet (2000) assesses different forms of mandatory risk pooling arrangements across insurers in cases of inadequate risk-adjustment (i.e. if the premium-replacing payments do not reflect risk and induce insolvency and risk selection). He uses three different data sources, a random sample of 69,000 insured individuals from a Dutch fund between 1992-1993, a simulation data set of 100,000 insured and aggregated plan level data for several Dutch insurance funds. He compares the predictive performance of two high-risk and two high-cost pooling models. His results suggest that that mandatory high-risk or high-cost pooling schemes are useful supplements to the current imperfect capitation formula for reducing selection and solvency problems.

Van Barneveld et al. (2001)[55] describe several forms of risk sharing between insurers and the regulator in competitive health insurance markets with imperfectly adjusted capitation payments. Optimal risk sharing depends on the weight regulators assign to reducing risk selection versus retaining efficiency. This article uses data for six consecutive years (1988-19993) for about 47,000 individuals enrolled in one Dutch sickness fund. In an empirical analysis they simulate the effects of four forms of risk sharing as a

[55] Like Schokkaert et al. (1998), this article present an empirical analysis of trade-offs between selection and efficiency.

supplement to demographic payment systems. The empirical results strongly support the conclusion that risk sharing for high-risks and high-costs yield better trade-offs between selection and efficiency than outlier or proportional risk sharing.

The majority of the discussed studies using the fourth method approach suggest selection evidence, mostly in the form of favourable or unfavourable risk selection (14 out of 23). Some conclude adverse selection or favourable selection (6 out of 23). Three publications in the list of 23 reported ambiguous evidence for selection or their results were otherwise not interpretable as selection evidence.

Table 2-5, on the next page, summarises all of the identified methodological approaches for empirical investigations into selection processes as discussed in the previous paragraphs. The following paragraphs of this section will discuss some of their shortcomings and problems.

Problems with previous empirical approaches: Moral hazard and others. The problem with the first approach is that it is hard to deduct from only health insurance coverage determinants, such as socio-economic and health status variables, whether adverse selection or risk selection occurs. While socio-economic and health status characteristics are valid general health insurance demand indicators, they do not necessarily translate into assumed (but unobserved) utilisation patterns that might indicate selection processes. Also, analysis results using this approach suggest that the choice between public and private health insurance coverage underlies tight affordability constraints and thus might be more a response to price and other factors.

A familiar problem of approach two is that adverse selection effects can be empirically clouded by another effect that also takes place within the context of information asymmetry: moral hazard. The two processes are impossible to distinguish in simple cross-sectional data.[56]

Under adverse selection high risks are much more likely first, to choose a complete coverage contract and second, to have higher health care utilisation. While moral hazard has the same observables, here the causality is different. Households obtain insurance contracts with coverage that exceeds their actual risks[57] and these households have less incentive to be cautious.

[56] This problem is widely acknowledged in applied contract theory, where one can also find the most recent approaches to overcome this issue, see for example Abbring et al. (2003).

[57] A household might either choose to be overinsured because of its high risk-aversion, or it might be bound to a compulsory risk pooling system, for example compulsory employment based health insurance.

Table 2-5: Methodology approaches for the empirical identification of selection processes

Type/Name	Type 1 Multiple-plan choice approach w/out utilisation	Type 2 Multiple-plan choice approach with utilisation	Type 3 Multiple-plan analysis approach	Type 4 Risk-adjustment approach
Analysis aim	Explain differences in insurance coverage with set of socio-economic and health status variables with biased selection	Explain differences in insurance coverage with set of utilisation variables (sometimes including socio-economic and health status variables) with biased selection	Explain differences in plan membership with plan characteristics with biased selection	Develop/validate predictors of individual health expenditures in order to risk-adjust and counteract selection
Analytical approach	Compare individuals with private vs. public insurance or insured of distinct insurance plans	Compare individuals with private vs. public insurance or insured of distinct insurance plans	Compare distinct plans	Assess explain-able variance of different expenditure determinants
Analysis attributes and dimension	Determinants of different insurance coverage: • Socio-economic • Health Status at one point in time or at different time periods	Determinants of different insurance coverage: • Health care utilisation and others at one point in time or at different time periods	Plan characteristics: • Price or Cost • Size • Risk profile etc. at one point in time or at different time periods	Determinants of insured: • Socio-economic • Health status • Health care utilisation and diagnostic information
Analysis level	Individual	Individual	Aggregated (plan level)	Mostly individual
Data	Cross-sectional data, pooled cross-sectional time-series data or two-period cross-sectional data	Cross-sectional data, pooled cross-sectional time-series data or two-period cross-sectional data	Cross-sectional data, pooled cross-sectional time-series data, two-period cross-sectional data	Cross-sectional data, pooled cross-sectional time-series data
Analysis results	Evidence for adverse selection	Evidence for adverse and risk selection	Evidence for adverse and risk selection	Evidence for risk selection
Publication examples	• Browne (1992) • Cutler and Reber (1998)	• Sapelli and Torche (2001) • Hopkins and Kidd (1996)	• Price et al. (1983) • Soderlund and Hansl (2000)	• Newhouse et al. (1989) • Van Vliet and v. de Ven (1992)

The state of being insured actually changes the household's behaviour and leads to the same observation of higher health care utilisation. Thus in both cases it is observable and testable that households with better, more comprehensive coverage have higher health care utilisation and health benefits.[58] However, this correlation is not informative as to the direction of the causality, which makes the two processes indistinguishable in easily accessible cross-sectional data (see Abbring et al. 2003a). The papers of Abbring et al. (2003a, 2003b) summarise a research program, proposing that the dynamic (i.e. time-related) aspect of insurance relationships can help to distinguish adverse selection and

[58] The methodological approach type two compares health care utilisation patterns of either insured to uninsured households or of households with high coverage to households with low insurance coverage. If the higher insured group has higher utilisation levels this is seen as a measure of adverse selection (or moral hazard).

risk selection.[59] Panel data that follow individuals over several time periods, rather than simple cross-sectional data can help to distinguish moral hazard from adverse selection.

Another problem arises for the multiple plan analysis approach type three. Adverse selection and risk selection might occur simultaneously and are empirically hard to distinguish in aggregated data. For example, one observes the average risk profile of two plans, while presumably the first risk pool is subject to selection but the second is not. However, if the first plan experiences risk selection where high risks are dumped *and* also low risks leave because of adverse selection, its average observed risk pool would be shown as stable. Even if there would be anecdotal evidence for selection processes, no empirical evidence could confirm this.

Finally, the fourth risk-adjustment approach has some problematic limitations. As van de Ven et al. (1998) points out, the explained variance of individual health care expenditures with the risk adjusters is quite small if not marginal. Furthermore, this approach is limited to capitation systems and the aim of empirical publications of this type is ultimately to design fairer capitation formulas that prevent or reduce profits from risk selection. Adverse selection is not considered here *per se*.

2.3.4 Summary and discussion of the literature review

This chapter's selective literature review can be by no means seen as a comprehensive account of all literature on selection processes (which would far exceed the limits of this thesis). However, the theoretical concepts of the selection processes were discussed and the state of empirical evidence for selection processes thoroughly assessed.

The literature focuses on adverse selection problems rather than risk selection problems. This is because adverse selection has the more obvious effect of market inefficiency, while risk selection has the more obvious effect of inequity. Thus, the mainstream research interest seems to be efficiency and not equity. This might be related to a number of factors. First, selection processes are predominantly discussed in the realm of economics and theories of market equilibrium where efficiency problems are of central concern. Second, the largely Anglophone literature reflects the dominance of a more liberal approach towards the economy, markets and the role of the state. Here one can observe a clear division of responsibilities between market and government. The

[59] Chiaporri and Salanie (2003) provide a survey of empirical applications to contract theory with several attempts that empirically investigate insurance markets for effects of asymmetry in information, like adverse selection and moral hazard.

government's role is to pitch in when unacceptable equity concerns arise as a market result or due to a market failure. In all other cases the market is generally seen as the most efficient device for the allocation of goods and services (Bozeman and Bretschneider 1986). Or as Pauly and Herring 1999 put it, these (equity) concerns are, "much more a matter of opinion," (p. 10), of judgement about equity and of relative political power[60] (see also Pauly 2002).

Although the literature tends to ignore risk selection or discusses it in the context of risk-adjustment and capitation systems used by high-income countries' national health insurance systems, the issue of risk selection becomes much more significant when addressing low- and middle-income countries. This is because low- and middle-income countries have fewer resources with which to support public health care systems. Privately financed and/or provided health care often might be the only accessible health care in these countries. This means that individuals who cannot afford private health insurance risk going untreated in the event of ill-health. Risk selection further undermines the ability to receive health care of those most likely to be ill and least likely to be able to afford private health insurance. Thus risk selection creates an equity issue of a different dimension in low- and middle-income countries. While in high-income countries risk selection causes the sick and poor to receive inferior care, in low- and middle-income countries the sick and poor will receive no care whatsoever. It is important to consider selection processes in low- and middle-income countries, because their practical outcomes diverge from outcomes in high-income countries.

The review of the empirical evidence for adverse selection and risk selection led to the following conclusions. The first conclusion is that there is a general lack of empirical studies of selection processes for low- and middle-income countries. It would be, therefore, an important extension of the current literature to conduct such a study - considering that selection processes might have an impact of different significance on health insurance markets in low- and middle-income countries than in high-income countries. Although the methodological challenges of empirical studies are similar in high or low- and middle-income countries, an empirical study in the later setting may improve understanding of the causes of inequity and inefficiency in low- and middle-income countries' health insurance markets.

[60] Equity concerns in case of health insurance risk pooling failure is associated with non-economic risk factors. Transfers in these cases are matters of judgement about equity and of relative political power.

Second, a majority of the reviewed empirical studies, applying a variety of methodological approaches to different data, concluded adverse selection and/or risk selection to some extent. Finally, there was a range of different methodologies applied in these empirical studies. The general challenge for all these methodological approaches consists of the fact that the study of motivations for observed individual behavioural outcomes has practical and statistical limitations. Thus most of the applied methodologies are based on rather intuitive approaches towards selection processes and attempt to derive statistical inferences from simple regression models applied to cross-sectional data or pooled cross-sectional time-series data. Two favoured attempts that cater to specific data measures can be generalised.

1. A favoured approach for the identification of adverse selection is to compare different plans or to compare individual plan choices under the assumption that more generous health insurers become subject to adverse selection. Generosity is measured here in terms of extent of coverage, premium price, institutional arrangements that influence access to health care and ultimately health care utilisation. This approach takes advantage of the fact that selection processes are grouping phenomena, i.e. some plans end up with disproportionate numbers of high risks and others with high proportions of low risks. High risks are characterised as individuals with certain socio-economic characteristics, like old age or geographic area of comparably greater poverty, poor reported health status and high prior health care utilisation. A potential problem with this approach is that observations attributed to adverse selection might be clouded by another process, moral hazard.
2. The favoured approach for the identification of risk selection is to analyse plans' risk pools under the assumption that insurers have an incentive to favourably select low risks (in order to operate more efficiently or maximise their profit). This approach is based on the belief that individual health care expenditures can be explained with a set of variables combined into individual risk profiles. The risk profile includes socio-economic variables, health status variables and prior utilisation variables. However, this body of literature is devoted to a rather different empirical problem than that of simply identifying adverse selection and risk selection in unregulated health insurance markets. It tries to identify practical methods that would eliminate the problem of cream-skimming (while at the same time preserving incentives for efficiency) in competitive health insurance markets with risk-adjustment/capitation systems.

The presented review revealed gaps and challenges. First, selection processes in health insurance markets are understudied in the context of low- and middle-income countries. This is of particular concern because of the specific impact these processes have on equity and efficiency in those settings. Second, previous methodological approaches of empirical investigations into selection processes faced considerable challenges, for instance distinguishing selection processes from other health insurance market problems like moral hazard. These distinct methodological challenges most likely originated from the fact that the well-developed theory on selection processes does not lend itself very well to empirical data testing. Additionally there is a shortage of easyly available data. Given this, the objective of the next section is to develop the research question and research objectives, which will address these identified gaps and challenges.

2.4 Development of the thesis research question and research objectives

The purpose of this section is to develop this thesis' research question and research objectives based on the literature gaps identified in the previous section. First, the research question is developed and the South African case study used to examine this research question is briefly introduced. Second, the research objectives are described.

2.4.1 Research question development and case study background

From the presented literature review it could be concluded that there is a very well developed theory on adverse selection and risk selection in health insurance markets. The literature recognises that the problems of adverse selection and risk selection either create inefficiency or inequity. However, a gap in the literature regarding the study of selection processes in the context of low- and middle-income countries was identified. Selection processes in those settings should be studied because of their diverging practical efficiency and equity outcomes.

This thesis' research question is: Are *unregulated* health insurance markets characterised by adverse selection and/or risk selection, thereby creating inefficiency or inequity? It was pointed out that unregulated health insurance markets either emerge from health insurance arrangements, which were never subject to government regulation or they emerge from arrangements, which were initially regulated and became subject to less regulation, i.e. deregulation. An interesting case for a middle-income country is the

unregulated health insurance market of South Africa's medical scheme plans in the context of its post-deregulation experience of 1995-1998, after premia risk-rating was legalised. (A general case study background can be found in Appendix 1.)

South Africa's medical schemes started to develop more than one hundred years ago as company-based, quasi-compulsory health insurance plans (Soderlund et al. 1998). The first medical scheme was the De Beers Consolidated Mines Ltd. Mines Benefit Society established in 1889 (Soderlund and Hansl 1999).[61] These schemes were all mutual societies, generally employment-based, and regulated by the government as what were called "Friendly Societies" - a broad term used to cover mutual insurance, savings and pension arrangements.

It was only in 1967 that legislation was promulgated, in the form of the Medical Schemes Act, recognising medical mutual insurers as distinct entities and establishing two bodies, the Council for Medical Schemes, and the Registrar of Medical Schemes, to fulfil the executive functions of the Act. Between 1969 and the mid 1980's government involvement in the medical schemes industry, via the Act, and various revisions thereof, was considerable. Regulation concerned the relationship between medical schemes and the provider community, and both the modes and rates of reimbursement were fixed by statute. Schemes were forbidden from charging differential premiums based on risk of ill health and were required by law to cover a certain percentage of the nationally mandated fee schedule for all health care provided.

In the following years, however, the market joined a worldwide trend towards economic liberalisation as people became disillusioned with centrally planned economic systems and the efficacy of government intervention in general (Peltzman 1989). In the 1989 amendment to the Act, rules disallowing risk-rating of premiums and the stipulation of minimum reimbursement rates were removed from the statute books, thereby making the sick and elderly much more vulnerable to losing medical scheme coverage (Soderlund and Hansl 2000). After the abolition of the apartheid system, the Medical Schemes Amendment Act (1993) turned the medical scheme system into a quasi-private industry. Medical schemes were now able to compete with medical insurers as they were allowed to offer different packages, to underwrite and to reinsure (McIntyre and Bowen 1994). A further deregulation in 1994 did away with the regulation of reimbursement rates altogether, and allowed contracting and vertical integration between providers and

[61] By 1910 there were seven such schemes in existence, and by the beginning of the Second World War in 1939, there was a total of 48.

medical schemes, paving the way for managed care arrangements. By the end of 1997, 176 medical schemes had been registered under the terms of the provisions of the National Health Act, 1977 (Soderlund and Hansl 1999).[62]

The South African medical schemes are divided into so-called 'open and 'closed' funds. Closed schemes were purely employer-based, mutual, non-profit organisations with community rated premiums. Membership in these funds was mostly compulsory for employees. In the past risk selection and adverse selection tended not to be real problems here. Open schemes were free for public enrolment. As designed, it would be much easier for open schemes to differentiate their premiums according to risk and thus practise risk selection. These schemes might also have been subject to adverse selection. Although open and closed schemes are non-profit organisations by law, it is generally considered that they function as for-profit enterprises in that the administration fees largely constitute the fund surplus. Also, professional administration companies administer both open and closed schemes. The processing of claims and the collection of premiums is mostly contracted out to these for-profit administrators, who are supposed to act on the instructions of the scheme members. In fact, with these administrators medical schemes can be seen as quasi-private health insurers.

The Health Systems Trust (HST) (1996) expected early on that the deregulation law, which allowed risk-rated premia, would reduce cross-subsidisation within medical schemes. Also, medical schemes were expected to become increasingly unaffordable to the elderly, chronically ill and low-income earners who would now have to rely more heavily on public sector health services (HST 1996). Indeed, several reports indicate unbalanced risk concentration in South Africa's unregulated medical scheme plans due to adverse selection and risk selection (for example: Soderlund and Hansl 1999, Soderlund and Hansl 2000).

During the years immediately following the deregulation measure, 1989 and onwards, medical scheme membership declined (Rama and McLeod 2001). This happened despite the fact that there are tax-concessions for health insurance premiums[63] (Price et al. 1996) and that a considerable amount of government finances goes to medical schemes through subsidies for civil servants. These subsidies are much higher than comparable

[62] More than 30 industrial medical schemes operated under the terms of agreements drawn up with the Labour Relations Act, 1956 (Act 28 of 1956).

[63] The South African income tax exempts the employer medical scheme contribution from the tax base (which can be up to two thirds of the employees contribution).

government health funding for the poor.[64] In addition to the observation that medical scheme membership declined it was also noticed that medical scheme contributions and benefits rose rapidly (South African Republic 1994).[65] However, the contribution increases that were observed for many medical schemes did not keep up with increases in expenditures on benefits (McIntyre 1993) and many medical schemes faced serious financial/solvency problems (South African Republic 1994).[66] The absence of substantial cost-containment efforts raised serious efficiency concerns.

The rapid increase in contributions prompted many young and healthy members to seek alternative insurance coverage with for-profit insurers or to give up insurance coverage altogether (HTS 1996). Thus medical scheme membership became increasingly unaffordable to many South Africans and the envisioned expansion of the medical scheme market to low income earners did not occur. The South African government recognised the problem that private medical insurance coverage became increasingly unaffordable for many South Africans. Subjects who were able previously to purchase private health insurance were driven out of medical schemes to rely on free or subsidised care in the public sector. The Department of Health (DoH) (2001) noted that equity could be served by retaining medical scheme clientele in the private sector. The argument was here that these households should not rely on the public sector, because this would diminish the limited resources for the poor and sick.

2.4.2 Research objective development

The literature review revealed a gap in the literature for *empirical* studies of adverse selection and risk selection, and particularly for low- and middle-income countries. Distinct methodological challenges have likely inhibited this research, because the well-developed theory on these selection processes does not lend itself very well to empirical data testing. Moreover, empirical studies in low- and middle-income settings often face serious data limitations.

Given these gaps in the literature, the research objective of this thesis is to conduct an empirical investigation that delivers evidence of adverse selection and risk selection in unregulated health insurance markets. The intention is to present this empirical evidence

[64] Up until now only about half of the eligible civil servants have chosen medical scheme coverage.
[65] Medical scheme membership, for instance, declined by nearly 4% between 1991 and 1992. Medical scheme contributions were equivalent to 7.1% of average formal sector salaries in 1982. Ten years later, in 1992, they amounted to 15.2% of average salaries (South African Republic 1994).

with an empirical method from which testable hypotheses can be derived. Also, in view of practicability and applicability to low- and middle-income country settings this method should be easy to apply to basic insurance data and standard software.

The first research aim is to derive an appropriate method applicable to empirical investigations into adverse selection and risk selection with available quantitative techniques and easily accessible data. The second research aim is to apply this method to an empirical analysis of case study data of an unregulated health insurance market in a middle-income country, here panel data of insured households in South Africa's medical scheme plans, 1995-1998. However, it will be not the goal of this thesis to advance the discussion of how to best distinguish adverse selection from moral hazard for the following three reasons. First, this thesis' data do not lend themselves to this purpose. The panel data under investigation is too short (four years) to really be able to distinguish these two processes. Also, the methodological approach that will be developed in the next chapter will allow the allocation of most of the analysis results to adverse selection rather than moral hazard.

Second, distinguishing between adverse selection and moral hazard is rather of a theoretical nature and will not be of main importance for the purpose of this thesis. Moral hazard was basically defined as over-utilisation, i.e. utilisation top-up. This does not usually concern the same type of utilisation as in the typical cases of unpredicted ill-health. Although most households who do practice moral hazard might have had some episodes of ill-health and subsequent health care utilisation, they simply extend, i.e. top-up their utilisation unnecessarily. Moral hazard effects are likely to be strongest for extra diagnostic procedures, such as seeing a (or another) specialist in order to confirm a diagnosis. They might also be found in the more frequent utilisation of discretionary services, like physiotherapy and optical services. In health care systems where public and private health care provision of different quality exists, moral hazard might be manifested in the choice of private over public hospitalisation.

However, if households end up with comparably high utilisation of any of these services it would not really matter for an insurance company if this was because this household is indeed a high risk or it just behaves like a high risk. This household is *de facto* rightly classifiable as high risk, no matter whether high claims/benefits are the result of adverse selection or moral hazard. Also, this thesis' data indicates considerable cost-

[66] Medical schemes operate on a pay-as-you-go or cash-flow basis, i.e. benefit payments are largely from contributions received during that same year (Prama and McLeod 2001).

sharing, which diminishes the incentives to practice moral hazard. Nevertheless the empirical analysis will distinguish between different types of health care utilisation that are more likely among low risks as compared to high risks. For example, evidence for high utilisation of discretionary services will not be seen as evidence for high risks and adverse selection. It will be interpreted as a type of top-up utilisation by low risks most likely initiated by favourable risk selection strategies (i.e. plan manipulation). High utilisation of basic/emergency services will be interpreted as evidence for high risks and possibly adverse selection. Also, in each model the types of health care services utilised will distinguish between public and private hospital utilisation.

Finally, from the literature review, and in particular from previous empirical studies, it is clear that the major problems in unregulated health insurance markets are selection problems. Previous empirical studies could somewhat identify selection processes and most publications assume that higher utilisation is related to adverse selection by high risks and not related to moral hazard of low risks. In particular, the few publications of the multiple-plan-choice approach that claimed to observe moral hazard draw contradicting conclusions from similar results. While Savage and Wright (2003) conclude moral hazard from the observations of longer hospital-stays, other studies, such as Sapelli and Vial (2003), could not find evidence for over-utilisation in the case of hospitalisation, for either public insurance or private insurance (since the price elasticity of demand for hospitalisation is low).

2.5 Summary

This chapter provided the analytical context for the thesis' research question and developed the rationale for this research question, which asks whether the two selection processes of adverse selection and risk selection increase inefficiency and inequity in unregulated health insurance markets. The research question was developed based on the literature review and the identified literature gaps, which this thesis wishes to address.

The second part of this chapter introduced the economic model of human behaviour as the analytical context of the thesis. Its underlying behavioural assumptions were used to explain the two selection processes. Behaviour (like adverse selection) of the insured is motivated by economic incentives and poses a serious challenge to health insurers. It can be argued that this behaviour forces insurers into practices (like risk selection), which are inconsistent with health policy objectives or do not allow coverage

where really needed - i.e. they fail to fulfil their actual function (Berliner 1982, 1984). However, because the effects of adverse selection and risk selection are unsatisfactory for all markets participants, it is predictable that individuals will have an incentive to protect their health insurance system with institutions other than the market, such as state-regulation.

The third part presented a review of the literature on adverse selection and risk selection in health insurance markets. The critical review of empirical studies that investigate selection processes identified several methodological approaches, ranging from multiple plan choice to risk adjustment methods. From the literature review it was concluded that while there is a large and comprehensive body of literature discussing theoretical aspects of selection processes, literature dedicated to the empirical identification of selection processes in data is less developed. Also, most literature is concerned with the description, discussion and analysis of health insurance problems in high-income countries. The situation of health insurance arrangements in low or middle-income countries is understudied, particularly in regard to empirical investigations into selection processes.

Finally, the fourth part developed the thesis' research question and research objective, based on the identified literature gaps. This thesis will contribute to the literature that discusses and empirically analyses adverse selection and risk selection in unregulated health insurance markets in low- and middle-income countries. The research question of whether unregulated health insurance markets are characterised by adverse selection and/or risk selection, thereby creating inefficiency or inequity will be examined in the context of a middle-income case study - South Africa's medical schemes' post-deregulation experience, 1995-1998. This thesis's research objective - to conduct an empirical investigation that delivers evidence of adverse selection and risk selection - is justified by the result of the literature review, confirming a scarcity of empirical studies in low- and middle-income countries. The primary challenge and first research aim of this thesis will be to derive a methodological approach for the empirical analysis of selection processes with available quantitative techniques and easily accessible data. Herein will lay the first main contribution of this thesis. The second research aim is to apply this method to the empirical analysis of the South African case study data. This will be the second main contribution of the thesis, adding to the limited literature on empirical investigations into selection processes in low- and middle-income settings.

The next chapter will develop the method for the thesis' empirical investigation.

Chapter 3

Method for empirical testing of adverse selection and risk selection: a group approach

3.1 Introduction

The purpose of this chapter is to pursue the first research aim and derive a method from which empirically testable hypotheses for the presence of adverse selection *and* risk selection in health insurance markets can be derived. First, a group approach will be presented. The literature review showed that the standard concepts of adverse selection and risk selection are usually explained on a time-scale. They will be reformulated as grouping phenomena. Previous methodological approaches, discussed in chapter two, used other aspects of selection processes' group dimension. Second, the empirical strategy that applies the group approach will be discussed in regard to applicable quantitative techniques and data requirements. It was a stated research aim that the empirical method should be applicable to easily accessible data and to quantitative techniques that are available on common software. Finally, this chapter will develop the thesis' research hypotheses based on the developed group approach.

After this introduction, part two of the chapter will derive the empirical methodology of the group approach. The key assumption will be that adverse selection and risk selection can be formulated as grouping phenomena. Grouping here means that selection processes can be distinguished by a segmentation of the insured population into easily identifiable groups. The empirical group method is based on two theoretical hypotheses that first, unregulated health insurance markets experience inefficient risk

pooling due to selection processes, and second, that groups of people who take part in selection processes have specific observable risk profiles. These two hypotheses mostly were explored separately in earlier empirical approaches, namely the multiple plan-choice and risk adjustment literature discussed in the previous literature review. The multiple plan choice method looked at how people select themselves into plans of different generosity levels, while the risk adjustment literature developed methods of assessing individual risk profiles.

The third part will present the empirical strategy using the group approach. It will be shown that the group method lends itself very well to empirical testing with techniques that are available in most software packages, in particular discriminant analysis and logistic regression analysis. These techniques can explain group membership for subjects that belong to *a priori* defined groups using a multivariate risk profile. The risk profile can accommodate several independent variables, such as socio-economic characteristics, plan characteristics, and health status & health utilisation data that are expected to describe group differences. Empirically testable expectations can be formulated for variables of this risk profile. Data that contains this type of information typically comes in the form of either cross-sectional or panel data. This type of data is easily accessible from routinely collected individual insurance data or national health (status/expenditure) surveys.

Finally, the fourth part will develop the thesis' research hypotheses, based on the group method. The main research hypothesis for the thesis' empirical investigation will be that South Africa's unregulated medical schemes encountered a lack of market-wide risk pooling and a lack of risk pooling within the market (between insurance plan options). From this, three research subhypotheses can be formulated.

3.2 Method: a group approach

This thesis' first research aim is to derive a *method* that yields *empirically testable* hypotheses for adverse selection *and* risk selection. The purpose of this section is to derive a conceptually grounded approach that can be applied to empirical data. As shown in the review of the last chapter, adverse selection and risk selection processes are typically explained using the dimension of time, because selection processes span over several time-periods. However, this theoretical concept does not lend itself very well to empirical testing. Rather, methodological approaches in the literature, like the multiple plan choice

method, used some form of group comparison, e.g. comparing characteristics and health care utilisation between groups with different levels of insurance coverage. It is worthwhile to further explore the description of selection processes as grouping phenomena in order to explain the methodological approach for empirical investigations into evidence for adverse selection and risk selection. This section will first discuss adverse selection and risk selection as grouping phenomena, emphasising their group dimension as compared to their time dimension. Second, the group method will be derived, yielding empirically testable hypotheses for selection processes.

3.2.1 Adverse selection and risk selection as grouping phenomena

The concepts of adverse selection and risk selection are not defined to a level of precision typically required for empirical testing. This is because these concepts were developed as explanations for observations of failing or absent (insurance) markets that seemed economically irrational within simple supply and demand models. However, within models that do not assume perfect information and perfect markets the concepts of adverse selection and risk selection can explain these superficial irrationalities. Both concepts are based on assumptions of the economic model of behaviour and on assumptions of individual risk perception and risk behaviour. However, behavioural assumptions are not only hard to incorporate in economic models that describe complex market processes, but they are also difficult to verify empirically in real markets (as opposed to limited experimental settings). The heavy reliance of these concepts on behavioural assumptions makes it difficult to empirically test for selection processes as well as to distinguish them in data.

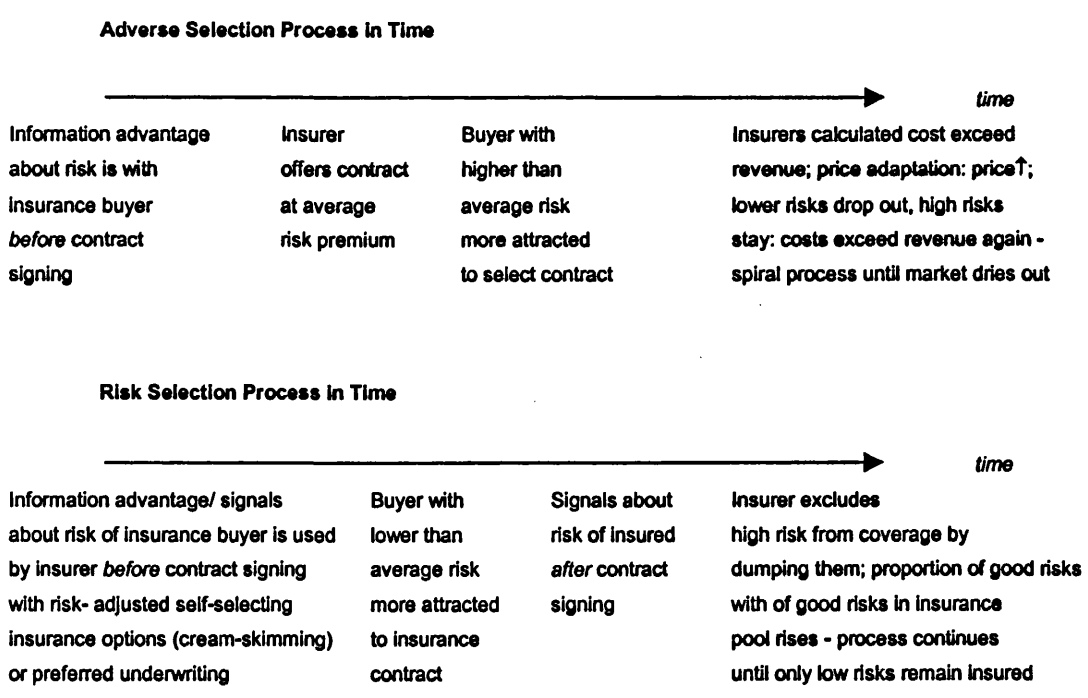
Delivery of empirical evidence for adverse selection and risk selection therefore requires the translation of standard definitions of adverse selection and risk selection processes into empirically testable classifications. This will be done through the description of adverse selection and risk selection as grouping phenomena. The following paragraphs will first compare the group dimension to the previously overemphasised time dimension when explaining adverse and risk selection processes. Second, they will suggest a typology for adverse selection and risk selection by looking at them as grouping phenomena.

A. Group dimension versus time dimension of selection processes

Using the time dimension is very useful for understanding selection processes conceptually and for illustrating their main features and consequences. However, the argument here is that the group dimension is, in addition to time, another very useful dimension of these selection processes.

By explaining selection processes using the time dimension, as in figure 3-1, it is theoretically possible to distinguish each step of each selection process for each individual case over time. But it fails to consider not only that both selection processes can fall together in time for any given subject, but also that several selection processes apply to different subjects and occur simultaneously. Empirically it is then difficult to distinguish between adverse and risk selection, since *both* processes overlap at any single point of observation for *different* observed subjects.

Figure 3-1: Selection processes time dimension

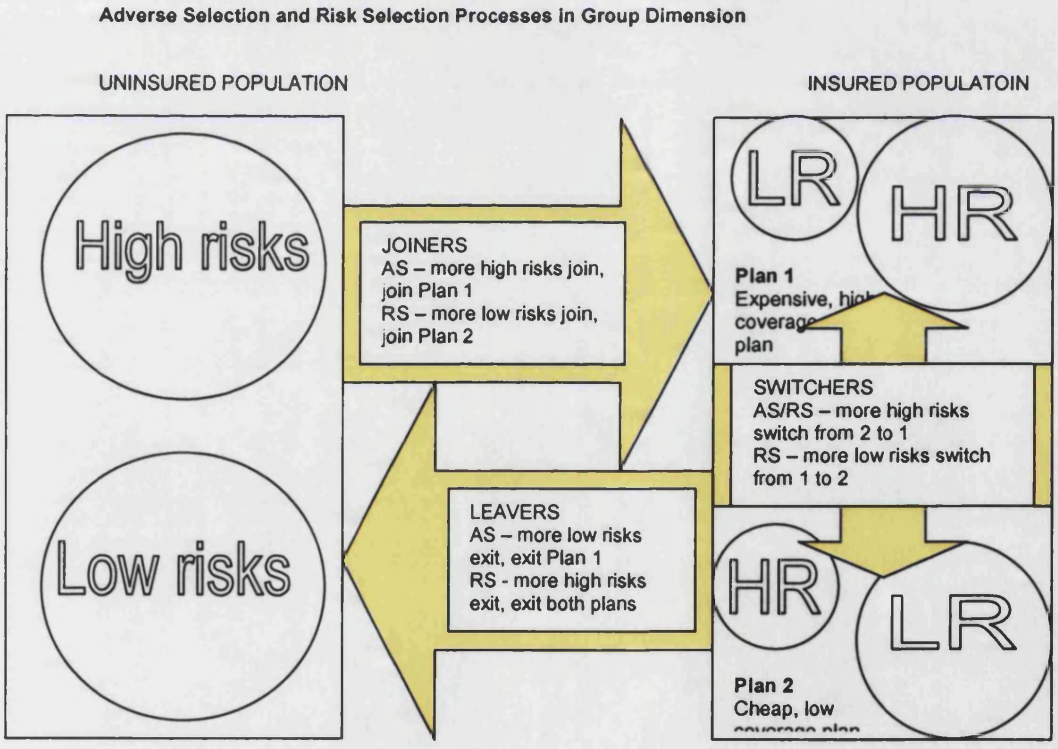


The predominant illustration of selection processes using the time dimension also limits the view on aspects that classify selection as grouping phenomena. For example, as a result of adverse selection it is assumed that insurance plans end up with a disproportionately high number of sickly high risks. Thus adverse selection is a process that concerns the movement of a specific group of people - those of high-risk - from the

uninsured population to the insured population where they select themselves adversely into specific plans - namely high risk plans with high-level, expensive coverage. Adverse selection also affects the group of already insured low-risks, who will move out of the insured population back to the uninsured population as a result of higher premium contributions.

As a result of risk selection, it is assumed that insurance plans end up with a disproportionately high number of healthy low risks. Thus risk selection is a process that concerns the movement of a specific group of people - those of low-risk - from the uninsured population to the insured population where they are risk selected into specific plans - namely low risk plans with low, inexpensive coverage fitting their needs. Risk selection also affects the group of already insured high risks, who will be forced to move out of the insured and back to the uninsured population. Figure 3-2 illustrates these processes of adverse selection (AS) and risk selection (RS) in the group dimension.

Figure 3-2: Selection processes group dimension



Note: High risks are abbreviated with HR and low risks with LR. Adverse selection is abbreviated with AS and risk selection with RS.

By explaining selection processes using the group dimension, as in figure 3-2, it is practically possible to distinguish adverse and risk selection processes in the market at any observation point in time. In a perfect risk-pooling market, subjects who belong to the groups of joiners, leavers or switchers are expected to have, on average, similar characteristics and risk profiles like the pool of the insured that they join (the control group). Thus, risks are pooled efficiently and there are average risk pools and average-risk plans. The uninsured and insured would simply sort themselves according to personal preferences that are not risk related.

However, the first key assumption here is that most markets are not perfect and for all these markets where selection processes are suspected, one could empirically observe that joiners, leavers, and switchers have, on average, different characteristics than the existing risk pool of the insured. The second key assumption is that imperfect markets also have inefficient risk pooling as a result of either adverse or risk selection. Thus in addition to average risk plans there are high-risk plans (with high coverage and high premium contributions) that attract high risks and there are low risk plans (with low coverage and low premium contributions) that attract low risks.

In summary, the explanation for, and illustration of, adverse selection and risk selection using the group dimension, as compared to the time dimension, seems an equally fitting and empirically more revealing approach. With the group dimension it is possible to define selection processes as grouping phenomena and derive methods that can assess whether or not subjects with certain characteristics, indicating their risk profiles, select into predicted groups. It is expected that this method, which considers this group dimension can lead to conclusive empirical evidence for selection processes in health insurance markets and make adverse selection and risk selection processes distinguishable from each other.

B. A typology for adverse selection and risk selection

A typology for adverse selection and risk selection is suggested here that is based on the group dimension rather than the time dimension. In order to illustrate this typology on the next pages two tables were developed. Different selection sub-types are distinguished in each column of table 3-1 for adverse selection and table 3-2 for risk selection. For each selection-subtype the table rows display: (1) manifestation (exclusion or attraction of low or high risks), (2) selection mechanisms, (3) effects, (4) affected membership groups, and (5) expected empirical observations if that type of selection takes place. Adverse selection will be approached as an issue of choice for certain risk types who want to become or stay

insured. The key assumption is that individuals' choices of insurance can convey information about their risk types. However, the main effect of adverse selection, the ultimate exit of low risks out of the market will be also considered as one type of adverse selection.

In table 3-1, on the next page, adverse selection type one in column one depicts the choice of insurance coverage by high risks. The manifestation is a high proportion of high risks entering insurance coverage. This mechanism can be described as adverse selection of high risks (possibly because of pre-knowledge about their poor health status). The effect is that high risks join insurance plans on average more often. Thus approaching this type of adverse selection as a grouping phenomena leads to the reasoning that the affected group is the group of joiners. The expected empirical observation is that, on average, joiners have characteristics that indicate a high-risk profile. They are, for instance, older and have higher claims than a control group of the already insured.

The typology also distinguishes two other types of adverse selection. Type two adverse selection is the choice by high risks to join certain insurance plans. This choice can manifest itself either by the entrance of high proportions of high risks into high-risk insurance coverage or the switch of high proportions of high risks to high-risk plans. High risk plans are typically plans that offer extensive or generous coverage at substantially higher prices. The affected groups where one could empirically observe adverse selection would be joiners and switchers. Type three adverse selection is the choice by low risks to discontinue insurance coverage in the observed market segment. This choice is manifested either by high proportions of low risks exiting insurance coverage or specifically, a high proportion of good risks exiting high-risk plans. The mechanism is here called adverse selection of low risks. However, this choice is likely to be induced by favourable risk selection, i.e. cream-skimming by competitors outside of the observed market segment - here called "external risk selection by competitors".[67] High-risk plans are exited by low risks, in particular, because they offer unnecessary coverage at too high a price for these risks. The effect is that, on average, more low risks exit high-risk plans. The affected groups that are of interest for exposing this type of adverse selection would be the leavers of insurance coverage.

[67] The result of the decision to discontinue insurance can be observed not its motivation. It might be that low risks are attracted to exit from the observed market segment as a result of favourable risk selection by external competitors. Therefore, it should be noted for the later interpretation of results that, although this type of observation is classified under adverse selection, it is probably linked to risk selection of external competitors.

Table 3-1: Adverse selection types as group phenomena

Adverse Selection	Type 1 **Choice of insurance coverage by high risks**	Type 2 **Choice of certain plans of insurance coverage by high risks**	Type 3 **Choice of discontinued insurance coverage by low risks**
MANIFESTATION	Entrance of High Risks	**A.** Entrance of High Risks into High-Risk Plans **B.** Switch of High Risks to High-Risk Plans	**A.** Exit of Low Risks **B.** Exit of Low Risks from High-Risk Plans
MECHANISMS	ADVERSE SELECTION OF HIGH RISKS	ADVERSE SELECTION OF HIGH RISKS	ADVERSE SELECTION OF LOW RISKS
EFFECTS	High Risks join on average more	**A.** High risks join on average more high-risk plans **B.** High risks switch on average more to high-risk plans	**A.** Low risks leave on average more **B.** Low risks leave on average more high-risk plans
AFFECTED GROUPS	JOINERS	**A.** JOINERS **B.** SWITCHERS	**A.** LEAVERS **B.** LEAVERS
EXPECTED EMPIRICAL OBSERVATIONS	On average joiners have characteristics that indicate a high-risk profile. (For example, joiners are on average older and have higher claims.)	A. On average joiners have characteristics that indicate a high-risk profile and join predominantly high-risk plans. (For example, joiners are on average older and have higher claims and join more often high-risk plans.) B. On average switchers have characteristics that indicate a high-risk profile and switch predominantly to high-risk plans. (For example, switchers are on average older and have higher claims and switch more often to high-risk plans.)	A. On average leavers have characteristics that indicate a low-risk profile. (For example, leavers are on average younger and have lower claims.) B On average leavers have characteristics that indicate a low-risk profile and leave predominantly high-risk plans. (For example, leavers are on average younger and have lower claims and leave from high-risk plans.)

Risk selection is approached by the insurer in terms of discouraging or attracting certain risk types, i.e. favourable selection and unfavourable selection of beneficiaries into health insurance coverage and certain health plans. Column one of table 3-2 depicts risk selection type one, discouraging high-risk insurance applicants. The manifestation is a low proportion of high risks entering the market. This mechanism can be described as unfavourable risk selection. Thus approaching risk selection as a grouping phenomena leads to the conclusion that the affected group is the group of joiners. The expected empirical observation is that on average joiners have characteristics that indicate a low-risk profile. The effect is that on average fewer high risks join. Joiners are, for example, younger and have lower claims than the control group.

Table 3-2: Risk selection types as group phenomena

Risk Selection	**Type 1** **Discourage high-risk insurance applicants**	**Type 2** **Attract low risks to certain plans of insurance coverage**	**Type 3** **Discourage high risks to continue insurance coverage or switch them to high-risk plans**
MANIFESTATION	Restricted Entrance of High Risks	**A.** Entrance of Low Risks into Low-Risk Plans **B.** Switch of Low Risks to Low-Risk Plans	**A.** Exit of High Risks (from High-Risk Plans **B.** Switch of High Risks to High-Risk Plans
MECHANISMS	UNFAVOURABLE RISK SELECTION	FAVOURABLE RISK SELECTION/CREAM-SKIMMING	A. DUMPING B. UNFAVOURABLE RISK SELECTION
EFFECTS	High risks join on average less	**A.** Low risks join on average more low-risk plans **B.** Low risks switch on average more to low-risk plans	**A.** High risks leave on average more (high-risk plans) **B.** High risks switch on average more to high-risk plans
AFFECTED GROUPS	JOINERS	**A.** JOINERS **B.** SWITCHERS	**A.** LEAVERS **B.** SWITCHERS
EXPECTED EMPIRICAL OBSERVATIONS	On average joiners have characteristics that indicate a low-risk profile. (For example, joiners are on average younger and have lower claims.)	A. On average joiners have characteristics that indicate a low-risk profile and join predominantly low-risk plans. (For example, joiners are on average younger and have lower claims and join more often low-risk plans.) B. On average switchers have characteristics that indicate a low-risk profile and switch predominantly to low-risk plans. (For example, switchers are on average younger and have lower claims and switch more often to low-risk plans)	A On average leavers have characteristics that indicate a high-risk profile and leave predominantly high-risk plans. (For example, leavers are on average older and have higher claims and leave more often high-risk plans.) B. On average switchers have characteristics that indicate a high-risk profile and switch predominantly to high-risk plans. (For example, switchers are on average older and have higher claims and switch more often to high-risk plans)

Two other types of risk selection are distinguished in the typology. Type two risk selection is favourable risk selection or cream skimming by attracting low risks to certain plans of insurance coverage. This choice can manifest itself either by the entrance of high proportions of low risks into low-risk insurance coverage or the switching of high proportions of low risks into low-risk plans. Low-risk plans are typically plans that offer limited/less generous coverage at substantially lower prices or plans that spend resources towards non-socially productive ends, like high discretionary benefits (Newhouse et al. 1997). The affected groups where one could empirically observe this type of risk selection

would be joiners and switchers. Type three risk selection is discouraging insured high risks to continue insurance coverage or to switch them into high-risk plans. This is manifested either by high proportions of high risks exiting insurance coverage or a high proportion of high risks switching to high-risk plans. In the first case this mechanism is described as dumping and in the second case as unfavourable risk selection.

However, one should keep in mind that risk selection is usually either a countermeasure to adverse selection by high risks or a measure to promote low risks to choose to become and stay insured. This fact leads to the problems of interrelation and overlapping between adverse selection and risk selection processes and the challenge of separating both empirically. Table 3-3 describes the overlaps in the expected empirical observations for the different adverse and risk selection types defined above. It specifies the affected groups that carry potentially similar information for the identification of either adverse selection or risk selection. The main overlapping problem is described as being between adverse selection type 2B (high risks chose to switch to high-risk plans), and risk selection type 3B (pressure on insured high risks to switch to high-risk plans).

Table 3-3: Overlap between expected empirical observations for different selection types

Adverse Selection / Risk Selection		Type 1	Type 2		Type 3	
			A. JOINERS	B. SWITCHERS	A. LEAVERS	B. LEAVERS
Type 1						
Type 2	A. JOINERS					
	B. SWITCHERS					
Type 3	A. LEAVERS					
	B. SWITCHERS			Switch of High Risks to High-Risk Plans		

From table 3-1 and table 3-2, above, it can be seen that the same characteristics are expected for switchers who switch to high-risk plans under adverse selection *or* risk selection. They are expected to have a high-risk profile. In the case of an empirical observation in which switchers display a high-risk profile and switch to high-risk plans it would be unclear which side (insured or insurer) initiated that process. Thus it could be that either adverse selection type 2B or risk selection type 3B took place. While adverse selection of type two cannot be clearly distinguished from risk selection of type three by

using information about the switcher group, data on other mover groups (joiners for adverse selection type two and leavers for risk selection type three) can still identify these selection types.

Another problem that is not exactly an overlap problem is pointed out for adverse selection type 3A (low risks chose to discontinue insurance coverage and/or switch to external competitors). The observed decision of low risks to leave might be due to the favourable risk selection efforts by external competitors outside of the observed market segment. In the case of an empirical observation with leavers displaying a low-risk profile it would be unclear which side (insured or external insurer) initiated that process. This problem, however, can be solved within the context of a specific observed market - for instance it has to be considered whether or not there is any information regarding aggressive recruiting by external competitors. Considering that the only overlap problem within this developed typology is the one for switchers it can be concluded that the generally impact of overlap problems is quite limited.

Based on this typology for adverse selection and risk selection, which considers the group dimension, the next section will derive the group method.

3.2.2 Group method

The group method will show that groups of people that take part in selection processes have specific observable risk profiles. The groups of interest can be defined by the nature of the selection processes. Selection processes also dictate the membership characteristics of these groups naturally. Even though there are different subtypes of adverse selection and risk selection (as discussed in the previous section) they all have two aspects in common. First, they are related to the enrolment of subjects into insurance coverage or the (voluntary or in-voluntary) disenrolment of subjects from insurance coverage. Only subjects who join, leave, or switch can either practice any selection or can be subject to selection. Second, they are related to the enrolment and disenrolment of subjects from and to specific insurance risk pools. Only subjects who join, leave, or switch between insurance plan risk pools can practice selection and establish insurance plans as a high- or low-risk pool.

First, the member group categories concern subjects that move into, out of, and within the observed health insurance market. These 'moving members', referred to from here forward as 'movers', carry valuable information regarding adverse selection and risk

selection in the observed insurance market. The aim of concentrating on different mover groups is to identify similarities and dissimilarities, which, in turn, makes it possible to conclude that either adverse or risk selection patterns exist. The primary interest of the group method is first, to form mover groups:

(1) Joiners, who newly enter health insurance coverage in the observed market;

(2) Leavers, who exit health insurance coverage in the observed market; and

(3) Switchers, between different plans within the observed market.

The group method then tests these groups of movers for significant differences in individuals' risk characteristics as compared to a fourth group:

(4) Control group members, who, insured in the observed market, never join, leave, or switch during the period of observation.[68]

For example, if subjects who recently joined insurance coverage in the observed market segment, on average, are more likely to generate claims compared to similar subjects in the control group, adverse selection could explain this observation. If recently joined subjects are, on average, less likely to claim benefits as compared to similar subjects in the control group, risk selection could be the explanation. Thus the group method would make adverse and risk selection processes in data visible and it would be possible to assess which of the two interrelated processes was prevailing during the period of observation.

Second, subjects will move to, from, and between plans of different risks within the observed health insurance market. Each insurance plan within the market can be defined as a risk pool. Different plans vary in their coverage and price level. Plans that offer generous coverage at considerable prices will most likely attract high-risk subjects and gradually turn into a high-risk pool. Movers' originating or destination plans provide valuable information regarding adverse selection and risk selection. The aim of considering the movers' originating or destination risk pools is to confirm expected adverse and risk selection patterns. It is of interest for the group method to classify plans into risk plan groups:

(1) Low-risk plans that, within the observed market, end up with a biased risk pool of predominantly low risks; and

(2) High-risk plans that, within the observed market, end up with a biased risk pool of predominantly high risks.

[68] Alternatively, depending on data availability, the control group could be also a group of uninsured subjects or subjects belonging to another insurance arrangement to which this market (segment) is compared.

The group method would then not only test the groups of movers for significant differences in individuals risk characteristics compared to the control group but also if they move to, from, or between low risk plans and high risk plans as compared to:

(3) Average-risk plans that, within the observed market, have an average risk pool.

For example, let us assume that the recently joined subjects, who were, on average, more likely to generate claims as compared to subjects in the control group also predominantly joined high-risk plans. Adverse selection, which was suspected above, can be supported with this observation. The alternative explanation of high risk-aversion can be excluded, since joiners had higher claims on average than the control group subjects. If recently joined subjects, who were, on average, less likely to claim benefits as compared to similar subjects in the control group, predominantly joined low-risk plans, risk selection could be confirmed. Thus the group method would help to attribute the existence of low- and high-risk plans to either adverse selection or risk selection.

The next section will discuss the empirical strategy using the group method. It will identify, in particular, quantitative techniques that allow the application of the group method to empirical data.

3.3 Empirical strategy with the group method: applicable quantitative techniques and data requirements

The first research aim of the thesis requires that the derived method has to be applicable to *available quantitative techniques* and *easily accessible data*. The purpose of this section is to discuss the empirical strategy using the group method, and, in particular, applicable quantitative techniques and data requirements. It will be shown that the group method lends itself very well to empirical testing with techniques that are available in most software packages. Also, the data requirements are that of routinely collected individual insurance data or national health (status/expenditure) surveys. For data that contain information on insurance status, socio-economic characteristics, and health care utilisation of individuals, specific variable expectations can be formulated. This section will first, discuss applicable quantitative techniques for the group method, namely discriminant analysis and logistic regression. Second, the data requirements for the group method will be discussed.

3.3.1 Applicable quantitative techniques: discriminant analysis and logistic regression

In order to apply the group method to empirical investigations into adverse selection and risk selection the method now has to be matched with applicable quantitative techniques. The group method assumes that subjects with certain risk characteristics (including originating or destination plan membership) are more prominent in identifiable joiner, leaver, and switcher groups as compared to a control group. The empirical research question becomes then: What are the characteristics and the plan membership of subjects that dominate mover groups as compared to the control group?

Thus applicable quantitative techniques have to be able to explain (and thus predict) group membership for subjects using a multivariate risk profile. The groups are defined *a priori* based on the group method. The risk profile has to accommodate several independent variables that are assumed to be able to describe group differences. In summary, the empirical technique needs to meet the following three technical objectives:

1. Evaluate group differences on a multivariate profile,
2. Validate the conceptual classification of observations into proposed groups, and
3. Identify and validate dimensions (variables) of discrimination between groups.

This limits the discussion to multivariate techniques, which assess the impact of a set of metric variables on a non-metric independent variable.

Multiple discriminant analysis allows a researcher to approach questions where the dependent variable is discrete (or categorical) and the independent variables are metric. Discriminant analysis explains the relationship that impacts the category in which an object is located. It is a profile analysis that can also be used as an analytical predictive technique. The main objective is to identify the group to which an object belongs and thus to understand group membership. It is desired that the group membership for each object can be explained or predicted by a set of selected independent variables (Hair et al. 1998). However, it can only identify group differences in cases where it is conceptually possible to define groups *a priori*.

Logistic regression is a special type of discriminant analysis confined to the testing of two independent categories or groups (Hair et al. 1998). Logistic regression has many advantages over classical discriminant analysis and is, for example, less vulnerable to the analytical assumptions of non-normality, non-linearity, and heteroskedasticity. It also allows the incorporation of dummy variables. Logistic regression generally delivers more robust results and has more standardised testing procedures (due to its similarity to linear

regression) that are widely available in different software packages (Hair et al. 1998). Logistic regression is, in particular, the better choice if a problem can be defined into two-group test problems making the interpretation of results easier as compared to cases with more than two groups (Hair et al. 1998, Hosmer and Lemeshow 2000). The group method showed that three two-group pairs could be formed. Appendix 2 provides a short introduction to the basic logit model and its assumptions.

In summary, logistic regression analysis (also referred to as logit analysis) - as a type of multiple discriminant analysis - is the best choice for the suggested group method. First, it is a multivariate technique that can study the relationship between a set of metric variables and a non-metric (discrete) independent variable. Second, with this technique all three technical objectives are addressed. It can evaluate group differences on a multivariate profile, it can test and validate the conceptual classification of observations into the *a priori* proposed groups, and it can identify the dimensions (variables) of discrimination between the groups.

3.3.2 Data requirements

Previously the group method could be matched with applicable quantitative techniques. Now the data requirements for applications of this group method have to be established.

The data needs to be on the individual level, either in the form of cross-sectional or preferably in the form of panel data. It should also allow the technical classification of observations into proposed groups and needs to identify and describe variables of discrimination between these proposed groups. Data typically collected by insurance companies or national health surveys contain the following required information about individuals: (1) insurance status, (2) socio-economic characteristics, (3) health status, (4) health care expenditures/utilisation, and (5) plan membership/characteristics.

For example, subjects that newly joined, recently left, or recently switched are usually distinguishable from the rest of the uninsured or insured population (i.e. members who were already or who are still insured) in data that records individual/household level insurance status changes. This information is in insurance files typically coded as the starting or end date of the insurance contract and in household surveys as the answer to the question regarding the household's insurance status (and more specifically if/when entered, exited, or changed insurance coverage). With this information it is possible to

identify the mover groups. Within the logistic regression framework this group classification would form the dependant variables.

Information regarding socio-economic characteristics, health care utilisation and plan characteristics are also part of routinely collected insurance data or household health surveys. For, example, plan membership and characteristics are distinguishable simply by the plan name, its organisational features, premium contributions or cost-sharing levels (that allow inference about its generosity or suspected biased risk pooling compared to average risk pooling). Socio-economic characteristics of individuals range from age, gender and income to residential information. Health status data relate to questions of self-reported medical history. Health care utilisation can be claim or benefit information from insurance data or expenditure information in household health surveys. Within a logistic regression framework these measures would be the independent variables. This linear combination of selected variables that will discriminate between the groups is also referred to as the risk profile of the observed subjects. The basic requirement of the risk profile is that it has to aid the distinction between low- and high- risk subjects. Independent variables in the risk profile have differing potential to contribute to the risk distinction. Each also has specifically formulated expectations under adverse selection and risk selection.

The risk adjustment literature typically uses the explained variance in annual individual expenditures to judge the risk discriminating potential of variables, like socio-demographic characteristics, health status information, and health utilisation in order to use them as risk adjusters (Newhouse 1996). It was reported in the literature review that socio-economic characteristics like age and gender are not very good risk adjusters, race and geography fare a bit better, but prior utilisation is the best risk adjuster. Self-reported health status information had mixed results as a risk adjuster, because it faces potential verification problems. Although socio-economic variables were identified as not very good risk predictors, they have to be included in potential risk profiles for statistical reasons (i.e. in order to control for their effects). Also, socio-economic variables are often the only 'observables' according to insurers sometimes risk select. Thus they might carry valuable information regarding selection processes (Brown et al. 1993).

If the socio-economic variable age would be included in a risk profile it could be expected that insurance coverage for the elderly is either harder to obtain or harder to keep, since the risk for ill-health rises with age. Thus it is likely that older people adversely select themselves into insurance coverage or that they are unfavourably risk

selected (i.e. dumped). Younger people are typically targeted by insurers' efforts to attract low risks. If joiners are older and leavers are younger than control group subjects then adverse selection can be suspected. However, as discussed before, if leavers are younger adverse selection might have been induced through the favourable risk selection efforts of other insurers. In this case it will be necessary to assess whether or not there is information regarding aggressive recruiting by external competitors. Adverse selection might also be operating if switchers are older than control group subjects, i.e. high risks chose to switch (to high coverage and thus high-risk plans). If joiners are younger and leavers are older as compared to the control group risk selection is likely. In the case of favourable risk selection switchers are expected to be younger than control group subjects (and switch to low-risk plans). Unfavourable risk selection is suspected if switchers are older (and switch to high-risk plans).

Variables that contain prior claim and benefits information are expected to be excellent risk indicators and factors for risk discrimination because they indicate whether or not a subject substantially utilised health services and is thus a low or high risk. Insurers always prefer subjects with prospective low risk, since they keep the costs low. Insurers are also eager to get rid of subjects with high claims and benefits. Under adverse selection joiners are expected to have higher claims and leavers lower claims as compared to the control group. However, low claims in the leavers groups might also indicate favourable risk selection by other competitors. Switchers are expected to have higher claims as compared to control group subjects under adverse selection of high risks (into high-risk plans). Under risk selection it can be expected that joiners will have lower claims and leavers higher claims than the control group. Switchers are expected to have lower claims as compared to the control group in the case of favourable risk selection (into low-risk plans). Unfavourable risk selection is suspected if switchers have comparably high claims (and switch to high-risk plans).

A plan classification variable that classifies plans with regard to their pooled risk would be another very useful indicator for risk discrimination. It is expected that subjects either self-select themselves or are selected by the insurer with regard to a plan's risk pool. Under adverse selection it can be assumed that joiners are more likely to be high risks and belong more often to high-risk plans and leavers are more likely low risks and belong more often to low-risk plans as compared to control group subjects. Adverse selection also might be operating if switchers are of high risk and belong more often to high-risk plans as compared to control group subjects, i.e. high risks chose to switch to high coverage and

thus high-risk plans. Under risk selection it can be expected that joiners are more likely low risks and belong more often to low-risk plans and leavers are more likely to be high risks and belong more often to high-risk plans as compared to the control group. In the case of favourable risk selection, switchers are expected to be of low risk and belong more often to low-risk plans as compared to control group subjects. Unfavourable risk selection is suspected if switchers are of high risk and switch to high-risk plans.

Based on this developed group method it will be possible to formulate this thesis' empirical testable research hypotheses. It will be developed in the next section.

3.4 Development of the thesis' research hypotheses

This thesis' main research hypothesis expects that one can observe a lack of risk pooling between high and low risk enrolees in South Africa's medical schemes. As discussed on the previous chapters, this does not exclude the possibility of a separating equilibrium, which might indeed be efficient but yet not equitable. This thesis's data and the empirical analysis design do not lend themselves to an explicit test for the existence of a separating equilibrium, however, this possible interpretation of the results will be discussed.

With the group method it is possible to derive empirically testable hypotheses from the thesis' main research hypothesis. First, this main research hypothesis can be formulated more precisely with the three subhypotheses shown on the left in figure 3-3. A lack of market-wide risk pooling might occur if, for example, most subjects, who recently joined insurance coverage, have had high-risk profiles. This might happen either due to classic adverse selection - as hypotheses 3A and B suggest - or induced adverse selection where external competitors risk select and attract all the low risks – as formulated in hypotheses 1A and B. A lack of risk pooling within the market might happen, if certain insurance plan options predominately pool either low or high risks. For example, some plans might be more aggressive in breaking risk pooling arrangements, like dumping high risks – as stated in hypothesis 2E - or by refusing coverage to high risks – as suggested by hypothesis 2A.

Second, with the developed selection process typology and the group method empirical hypotheses for each subhypothesis can be formulated, including the pair of membership groups that should be tested and the risk profile that can be expected. This is shown on the right side of figure 3-3.

Figure 3-3: Research hypotheses and corresponding empirical hypotheses

Research hypotheses	Empirical hypotheses
Hypothesis 1: There is competition for low risks between medical schemes and for-profit insurers. The previously regulated (observed) medical scheme market segment is contested, and experiences a disproportionate loss of low risks due to: A. Risk selection by (external) for-profit insurers for low risks in the medical scheme market segment, and (initiating) B. Adverse selection with low risks exiting medical schemes, and, in particular, exiting their high-risk plan options because insurance coverage here is comparably expensive and they can find better coverage at lower costs elsewhere.	Hypothesis 1: A disproportionate loss of low-risk members will be shown by testing the leavers and the control group for differences in their risk profiles. If medical schemes disproportionately loose low risks it will be possible to observe that: A. Under risk selection by (external for-profit insurers) leavers are, for example, younger and have lower claims than control group members; and (initiating) B. Under adverse selection, low risks' exit medical schemes, leavers are, for example, younger, have lower claims, and have more likely a high-risk plan membership than control group members.
Hypothesis 2: In this competition medical schemes follow several risk-management strategies to protect their scheme and plan risk pools including: A. Risk selection that discourages the entry of high-risk insurance applicants; B. Risk selection that attracts the entry of low risks with self-selection low-risk plans; C. Risk selection that prevents the exit of low risks by offering low risks to switch into attractive low-risk plan options; D. Risk selection in the form of pressuring high risks to switch into high-risk plan options; and E. Risk selection by discouraging high risks to continue their medical scheme coverage, particularly if they belong to high-risk plans.	Hypothesis 2: Med. schemes' risk selection strategies will be shown by testing joiners/switchers/leavers and the control group for differences in their risk profiles: A. Under risk selection that discourages the entry of high risks, joiners are, for example, younger and have lower claims than control group members; B. Under risk selection that attracts the entry of low risks, joiners are, for example, younger, have lower claims, and have more likely a low-risk plan membership than control group members; C. Under risk selection that prevents the exit of low risks, switchers are, for example, younger, have lower claims, and have more likely a low-risk plan membership than control group members; D. Under risk selection in the form of pressuring high risks to switch; switchers are, for example, older, have higher claims, and have more likely a high-risk plan membership than control group members; and E. Under risk selection by discouraging high risks to continue their medical scheme coverage, leavers are, for example, older, have higher claims, and have more likely a high-risk plan membership than control group members.
Hypothesis 3: High risks, who are either seeking medical scheme coverage or who are already insured in the previously regulated medical scheme market follow strategies of: A. Adverse selection by choosing to enter medical scheme plan coverage, particularly expensive and comprehensive high-risk plan options; and B. Adverse selection by choosing to switch into medical scheme plans with expensive and comprehensive coverage, typically high-risk plan options.	Hypothesis 3: Adverse selection strategies of high risks will be shown by testing joiners/switchers and the control group for differences in their risk profiles: A. Under adverse selection joiners are, for example, older, have higher claims, and more likely have a high-risk plan membership than control group members; and B. Under adverse selection switchers are, for example, older, have higher claims, and more likely have a high-risk plan membership than control group members.

For example, subhypothesis 3A, of adverse selection strategies of high risks seeking medical scheme coverage, in particular high-risk plan coverage, will be shown by testing

joiners and the control group for differences in their risk profiles. It can be expected that under this type of adverse selection joiners are expected to be older, have higher claims and more likely a high-risk plan membership than control group members. Under the type of adverse selection specified in subhypothesis 3B, where already insured high risks seek to switch to high-risk plan options, switchers and the control group will be tested for differences in the risk profile. Here it is expected that switchers are, for example, older, have higher claims and are more likely to have a high-risk plan membership.

3.5 Summary

This chapter presented the results for the thesis' first research aim - deriving an empirical method with testable hypotheses for adverse selection *and* risk selection, which is applicable to available quantitative techniques and easily accessible data. The group method incorporates elements from previous empirical methodologies, namely the multiple plan choice and the risk adjustment literature discussed in chapter two. The group method is based on the idea that concepts of adverse selection and risk selection can be better empirically tested by explaining them as grouping phenomena rather than as processes that span over several time-periods.

The first part of this chapter derived the group method. A typology for adverse selection and risk selection was developed, which approached the complexity of both selection processes by breaking them up into several empirical problems that are easier to test. The selection types were distinguished according to their manifestation, mechanisms, effects, affected groups, and expected empirical observations. Based on this typology the group method was presented, which can disclose in empirical studies whether or not subjects practised adverse selection or were risk selected, by observing whether subjects with certain risk profiles are more prominent in groups of movers and in certain plans. These mover groups are first (1), three different groups of movers - joiners, leavers, and switchers - as compared to a control group of non-movers, and second (2), groups of different plans - plans that pool low and high risks as compared to average risks. The group method supports the theoretical argument that adverse selection and risk selection are grouping phenomena and that empirical research should therefore apply methods that can assess whether or not subjects of certain risk really group into these specific predicted groups. From this approach empirically testable hypotheses can be derived.

The second part of this chapter identified applicable quantitative techniques applying the group method, and addressed the data requirements. Logistic regression was suggested as the best choice for empirical studies that want to apply the group method because it can explain group membership for individuals that belong to *a priori* defined groups using a multivariate risk profile. Thus the group method leads to three testable logistic regression models, which will yield probabilities that a subject or a population with certain risk profile characteristic belongs either to one of the three mover groups or the control group of non-movers. Significant differences in the specific risk profiles of subjects that are more or less likely to join, leave, or switch rather than to belong to the control group can then be explained by different selection processes. From the risk profile, data requirements for the empirical investigations that wish to apply the group method were derived. The data has to be on the individual or household level and in the form of cross-sectional or panel data. Either insurance companies or national household surveys with a health component typically collect this type of information. This data is relatively easily accessible.

Finally, the fourth part developed this thesis' research hypotheses and the corresponding empirical hypotheses based on the group method. This thesis' main research hypothesis, will be examined in the context of a middle-income country case study - South Africa's medical schemes plans.

The group method derived in this chapter will be the backbone of this thesis' empirical investigation into evidence for adverse selection and risk selection. The next chapter will apply this method to the South African case study data.

Chapter 4

Data & empirical analysis: the South African case study

4.1 Introduction

The purpose of this chapter is to pursue this thesis' second research aim by conducting an empirical analysis into adverse selection and risk selection applying the group method. The case study that will be used for this empirical investigation contains longitudinal data from the unregulated South African medical scheme market describing the time-period between 1995-1998. First, it is the aim to of this chapter to present the data that will be used for the empirical analysis. In the course of the data discussion, key issues that surfaced during the data examination and preparation will be described. In particular this concerns the data format, which is a longitudinal panel. Second, the empirical analysis and statistical models for this thesis will be presented. Logistic regression was identified in the last chapter as the best quantitative technique for the group method and the particular logit model (structure) for the data at hand needs to be specified. The software used for the data examination and empirical analysis is STATA™ versions 6 and 7.

After this introduction, part two of this chapter will present the data and data sample derivation for the empirical analysis. Discussed key data issues concern the identification and exclusion of outliers and missing values and the creation and transformation of crucial data variables in preparation of the analysis. The goal here is to provide the reader with an understanding of the data sample used for the subsequent analysis while also pointing out data peculiarities, namely its panel structure, that are essential for the following empirical model building strategy.

The third part will present the empirical strategy with a logit model for panel data. It will be shown that one applied logistic regression model for correlated data, the marginal or population averaged model, is the most appropriate logistic regression model for the thesis' empirical data analysis. The actual model building will follow a confirmatory model specification approach. This specification approach is used because the models will essentially be used to confirm existing theory and the group approach. The thesis' models are built on two hypotheses, one derived from the multiple plan choice literature and a second one based on the risk adjustment literature. With the identified model building strategy three models will be specified. These models will discriminate between data subgroups using a multivariate risk profile in order to derive empirical evidence for adverse selection and risk selection. Expectations for the independent variable vector that addresses this thesis' research hypotheses will be described and a four-step model building process will be applied.

4.2 Data and data sample: longitudinal panel for 1995-1998

This thesis' second research aim is to conduct an empirical analysis into adverse selection and risk selection with *data* describing an *unregulated health insurance market* in a *middle-income country*. The purpose of this section is to present and discuss the data that will be used for the empirical analysis. The data describes a sample of privately insured households in an unregulated market of a middle-income country, the Republic of South Africa. The part of South Africa's health insurance market that is of concern for this thesis is the medical scheme market, which has its historic origins in mutually organised employment-based insurance arrangements. Medical schemes were subject to strong government regulation between 1960 and 1990. The Medical Scheme Act (1967) made them subject to registration with the Office of the Registrar, recognising them as non-profit entities and preventing any risk-related premia-rating. Following a wave of deregulation in the early 1990s many medical schemes introduced risk-rated premia and turned into quasi-private health insurers.

Medical schemes cover 20% of the South African population and guarantee its member's access to good quality health care provided in the private sector. In 1998 total

medical scheme expenditure[69] compromised about 90% of all private health expenditure in South Africa (Rama and McLeod 2001). Medical schemes can be classified into so-called 'open' or free public individual/group enrolment schemes and 'closed' or restricted employment-, union- or other professional-based enrolment schemes. There were 160 to 180 registered schemes between 1990-1999, among which open schemes represented 30% and closed schemes 70%. During this time period the number of open schemes increased from 42 to 48 and the number of restricted schemes decreased from 138 to 112. The trend in membership development for the two different scheme types followed a similar pattern. Most registered schemes, about 60%, were small schemes with less than 6,000 members. However, the majority of open schemes were large, with more than 30,000 members, and the majority of closed schemes were small. Between 1990-1999 the number of all small schemes decreased from 138 to 89. The decline in the number of small registered schemes could be largely attributed to a decrease in small closed schemes (Rama and McLeod 2001). Medical schemes are either self-administered or professionally administered by for-profit companies (Soderlund and Hansl 1999). The data used here originated from a South African for-profit holding company that manages the administration of 29 medical insurance schemes.

This section will first discuss the data source and data sample derivation. Second, an overview of the data variables and study population will be provided. The overview will also comprise the description of how crucial identifier variables for the analysis were created. The later concerns a process of classifying plans into different risk pools and the depiction of the data subgroups through identifier variables. Finally, it will introduce the subgroups of joiners, leavers, switchers, and the control group and compare some of their household characteristics by taking the panel structure of the data into account.

4.2.1 Data source and data sample derivation

The data used in this analysis was collected by a holding company that administers several of South Africa's medical schemes. It became available in the course of a previous research project that investigated the effects of the deregulation policy between 1989-1994 on risk pooling of the medical schemes (Soderlund and Hansl 1999, 2000).[70] However, this

[69] Medical scheme expenditure accounted for 73% of South Africa's total private health expenditure. If the out-of-pocket expenditures of medical scheme members is included this number rises to 89%.

[70] The research project studied trends in risk pooling and efficiency of South Africa's medical schemes following deregulation. It was sponsored by the International Clearinghouse for Health Systems Reform Initiatives, the European Union (via the DoH), the Medical Research Council of South Africa, and the

data was never actually analysed prior to this thesis, because it challenged the capacities and time-constraints of that research project.

The background for this research project was the change in the regulatory regime for South Africa's medical schemes. The prior to 1989, under the Medical Scheme Act of 1967, regulated medical schemes became subject to a series of deregulations by the South African government between 1989 and 1993 (HST 1996, 1998, Soderlund and Hansl 2000). In particular the new rule that allowed the replacement of community-rating with risk-rating transformed the prior mutual medical scheme system into an unregulated quasi-private health insurance market with anecdotal evidence for adverse selection and risk selection. These anecdotal reports were recognised by the South African government and led to a research project in 1997-1998, which aimed to either verify or falsify these anecdotal reports with empirical evidence. The project, by Dr. N. Soderlund and the thesis author, analysed several data sets and conducted interviews and a document review in the office of the regulator, the Registrar of the South African Medical Schemes.

The holding company data, which are used in this thesis, represent the four-year time-period between 1995-1998, when most medical schemes switched to the quasi-private mode of operation. The last fact is based on the previous research project's observation that deregulation rules allowing risk-rating were only fully implemented by the medical schemes in the mid-1990s. This conclusion was drawn from interviews with the Registrar of Medical Schemes, Danie Kolver, his assistant Ryno van Zijl, and a document review of medical scheme files for the years 1989 to 1998 at the Registrar's office in Pretoria. These sources consistently indicated that most schemes did not implement risk-adjustment strategies, including premia risk-rating, until 1993-1995. In fact very few schemes started any risk-adjustment strategies by 1993, and most schemes took at least one or two more years before they established strategies that included premia risk-rating. Thus the thesis data of South Africa's unregulated medical schemes describe a post-deregulation experience. However, a before-and-after deregulation scenario for the analysis of this thesis' case study cannot be performed with the available data starting in 1995 (because some medical schemes might have adjusted to the deregulation of 1989 earlier than 1995 when the observed data starts).

The data provided by the medical scheme holding company contains a given set of 29 medical schemes. This represents a sample of about one-fifth of all South African

London School of Hygiene and Tropical Medicine (via the British Department for International Development).

medical schemes, which covered 20% of South Africa's population between 1995-1998. It is assumed that this can be seen as a representative sample of the South African medical scheme population. This is because the selection of medical schemes seems random, covering a variety of different schemes, former open-enrolment and closed-enrolment schemes and schemes of varying sizes (see table A4-1 in Appendix 4 and the description of these plans in subsection 4.2.2.A.). Although there is no founded indication, it might be that the selection of given medical schemes is not completely random, and the medical schemes administered by this holding company differ from other medical schemes in the market (producing biased results).

The data is in the form of a short (four-year) panel of insured individuals and their families (i.e. households) belonging to one of the medical schemes and one of their plan options. This type of data is often referred to as longitudinal data. The raw data compromises N=1,011,735 observations for n=353,458 insured households over the four year observation period between 1995 and 1998, where the unit of observation is the household level. The insured households are members of the 29 medical schemes with 49 plan options altogether. Each of the insured households has observations for at least one year and a maximum of four years. It provides detailed information on a series of households' socio-economic characteristics, levels of chosen coverage and premium contribution, actual claim and benefit patterns, and the dates when a chosen scheme and plan coverage started and ended.

Data is limited to what has been recorded and the data was originally not collected for the purpose of this analysis. For example, within the holding company data some variable measures were not complete for some household observations. As a result the data is partially incomplete - also called unbalanced. Missing data of this type are very common in longitudinal data sets, because repeated measurements for one subject/household have to be obtained but the circumstances for obtaining these measurements are often not within the control of the data collector. After the identification of outliers[71] and the evaluation of the missing data (process) it was justifiable to exclude the small percentage of cases with incomplete or outlier observations (12% or 177,980 of all observations, but only 8% or 27,501 of all households). Appendix 3 describes the diagnostic techniques used for the initial data examination, in the course of which the influence of missing values and outliers was assessed. Consequently the sample used in

[71] Compared to the majority of the sample, outliers are characterised by distinct differences. Outliers usually represent an extraordinary event that can either be explained by the researcher or not.

this thesis contains N=833,755 observations belonging to n=325,957 households over a four-year observation period from 1995 to 1998.

4.2.2 Data variables, study population, and created identifier variables

The data contains socio-economic characteristics, like age, gender, number of dependants, residential area, income and race for the so-called principal member. He or she is the individual that signed the insurance contract and is responsible for the premium payment and other occurring insurance costs. Most principal members in the initial sample (about two-thirds) have some type of dependant (spouses, children etc.).[72] This fact causes most data to reflect household values for more than one individual. Information that is reported for each household contains data on contributions, co-payments, claims and benefits. It would therefore be wrong to speak about insured individuals and their characteristics. In fact, the data mostly covers data per family or household per year, not per subject. Thus the unit of observation for the data analysis will be the household level.

The variables in the data were distinguished according to the following criteria. First, there are variables carrying socio-economic characteristics of the principal member, like age, gender, number of dependants, income group, race, and a geographical characteristic - the postcode of the residential area. Using the recorded post codes, it could be distinguished if the principal member's residential area was either rural or urban. This information was captured by creating a dummy variable (one if urban, zero otherwise). Second, several variables reflect claims, tariffs[73], contributions, and benefits per household. In addition, this second variable group included a range of different medical benefit variables for services like, GPs, specialists, public and private hospitals, dental, optical, and other services.

However, since this data was not collected for the purpose of this thesis, several of the original measures were converted into other more applicable (and for the empirical analysis, useful) variables. For example, all contribution-claim and benefit variables were converted into measures per household member and household membership month. More importantly, several new variables were created, namely the risk plan and mover group identifier variables, which are crucial for the analysis. The following paragraphs will briefly describe the data variables, including the plan classification process creating

[72] If the data indicates dependants these are covered by the insurance contract.
[73] In South Africa the tariff is the legally billable amount for medical services and procedures.

the risk plan variable and the determination of the data subgroups through identifier variables.

A. Data variables describing the study population

About two-thirds (70%) of the observed 325,957 households were headed by a male principal member. White South Africans constituted about a third of all household observations (107,672 or 33%), similar to the Coloured population (102,068 or 31%) and Black South Africans (89,852 or 28%). Asian South Africans were by far the smallest group of households with only 8%, or 26,373, headed by an Asian principal member.

Income was originally coded as an approximate income group category between zero and 999 based on the income when the household first joined. This coding was provided by the medical schemes holding company, the data source. It represents no actual monetary income value in Rand, but an income category assigned by the medical schemes. Zero was the lowest income group and 999 the highest income group. The mean income group was 549. The medical schemes were not willing to provide more accurate income data stating confidentiality reasons.

In this form the income variable was not very informative, and it was converted into a categorical variable, which takes the value zero if a household belongs to a defined average income group, comprising all households with income groups between 500-599. The variable has the value one if a household belongs to a below average income group between 0-499 and the value two if it belongs to an above average income group between 600-999. The income variable was coded into these three categories, average, below and above average, because it was hoped this would provide more easily interpretable results. The reference category will be defined as the average income group in the empirical analysis. Three percent of the data households had an average income. Thirty-nine percent of all household data belonged to the above average income group while 58% belong to the below average income group.

Table 4-1 presents the description of several continuous variables. The overall mean age (over all household-years) of the principal household member was 42 years and the average number of dependants was two. The average claims per member and membership month lay at 229 Rand, the average contributions at 242 Rand. A co-payment variable was created by calculating the difference between the legally billable amount (or

tariff) and the actual paid-out benefits.[74] The mean co-payment was 23 Rand. Among the benefits[75], private hospital benefits had the highest overall mean with 43 Rand per member and month, while public hospital benefits had the lowest overall mean with 2 Rand. Claim, co-payment, and benefit data showed a relatively high variation, which is due to the unpredictable nature of ill-health events.

Table 4-1: Description of selected continuous variables

Variable	Obs.	Mean	Std.Dev.	Min	Max
Socio-economic					
age	833,755	41.65	13.00	0	98
dep (dependants)	833,755	1.90	1.67	0	17
inc (income group)	833,755	548.82	341.01	0	999
Claim-contribution & benefits					
tac (claims)	833,755	228.81	440.38	0	94,471
tco (contributions)	833,755	241.61	121.50	0	8,600
co2 (co-payments)	833,755	23.33	123.58	0	44,554
gp (GP benefits)	833,755	22.86	21.04	0	1,457
pub (pub. hospital benefits)	833,755	2.02	49.82	0	31,480
spe (specialist benefits)	833,755	35.67	97.93	0	12,286
prv (priv. hospital benefits)	833,755	42.46	235.75	0	42,826
den (dental benefits)	833,755	16.53	37.81	0	2,503
opt (optical benefits)	833,755	10.68	21.62	0	997
par (paramedical benefits)	833,755	3.25	31.07	0	5,399

Note: This table does not take the panel data structure of the data into account and instead provides a rough summary statistic for these variables over all N=833,755 observations or household-years (rather than n=325,957 households). This is sufficient, however, in order to provide a basic data overview.

The observed households belonged to 29 different medical schemes that altogether have 49 plan options for their members. This sample of 29 medical schemes was determined by

[74] Of altogether three possible co-payment variables – (1) co-payment1 per person (claim-tariff), (2) co-payment2 per person (tariff-benefit), and (3) total co-payments per person (claim-benefits) - only the co-payment2 variable is considered here. The total co-payment variable is an aggregation of co-payment1 and 2. Both are two very different conceptual measures with opposing underlying trends over time. For this reason it seemed not to be a very useful variable for further consideration. Co-payment1 is a measure of provider-overbilling over the agreed legal tariff. It will not be considered further because it captures a form of cost-sharing, which could not be influenced by the insured household. This co-payment component would add no value to the analysis of selection processes. Also, the co-payment1 share of the total co-payment is significantly less (approximately 20-30%) than the one of co-payment2. The more significant part of the co-payment is contained in the co-payment2 variable.

[75] GP and public hospital benefits will be classified in this thesis as basic benefits, since both are seen as emergency or essential medical services. They represent two different service sectors – ambulatory and in-patient care. Specialist and private hospital benefits were classified as negotiable benefits. They are seen as negotiable, since cheaper options, like GP or public hospital services, exist. However, they are not discretionary. Medicine benefits, another available variable, was not considered further because it was highly correlated to other essential and negotiable benefits. A third set of the benefit variables was classified as discretionary benefits. Dental, optical and paramedical benefits will be considered further as they stem from optional and non-essential services. Two other available discretionary benefit variables of physiotherapy and associated/miscellaneous benefits are not considered further because the correlation analysis showed high

the data set taken from the holding company. It represents a typical sample of the medical scheme market, with some (9 out of the 29) schemes having had 'open' or free public individual/group enrolment, and other (20 out of 29) schemes having had 'closed' or restricted employment-, union- or professional-based enrolment.[76] A list of the medical schemes (open and closed), together with their overall and average number of observations over the four-year observation period is given in Appendix 4, table A4-1.

As in the overall number of the 160-180 schemes between 1990-1999 also this thesis' data sample of 29 schemes between 1994-1998 contained one-third open schemes. Applying the scheme size categories of the Office of the Registrar, small schemes having less than 6,000 members, medium schemes having 6,000-29,999 member and large schemes having more than 30,000 members, 21 out of the 29 schemes can be defined as small, six as medium and two as large. This distribution also coincides with the overall proportion of all 160-180 registered schemes classified as small or large. Among the 21 small schemes two were open schemes, the remaining were closed schemes. The number of open schemes was much higher for the medium and large category, seven out of eight. This too is confirming that the analysis sample is a good representation of the medical scheme market, where the majority of closed schemes was small and the majority of open schemes were large. This means of course that a small number of (open) schemes contribute the bulk of observations to the analysis. For example, the three largest open schemes contribute around 50% of all observations.

Since the data was collected and managed by one holding company, the data reflect standard reporting from the plans and can be assessed as reliable. Thus variation in the quality of data by scheme or plan is not expected and given the large data point number the quality can be assessed as good. No selection bias is suspected in terms of the analysed schemes, they represent a typical mix of open and closed, small and large schemes that either operated nation-wide or for specific industries. Altogether, the

correlation to all other types of benefits. Besides this, they seemed generally of lesser relevance, because of their low values.

[76] Official records at the Registrar of the Medical Schemes have only been classifying schemes as open or closed since 1997. However, the Registrar's office was internally classifying schemes according to this distinction since about 1992 (Rama and McLeod 2001). The schemes used in this sample were classified regarding open and closed with the help of the Registrar D. Kolver and his assistant R. van Zijl based on their knowledge and memory (in Summer 1998). However, if schemes changed from open to closed and vice versa during 1992-1997 their final classification was used here. Before 1995 open and closed schemes had roughly similar membership levels. However, in the period after 1995 membership of closed schemes declined while membership in open schemes rose. Also, the number of open schemes rose, while the number of closed schemes declined. This was largely due to a change in the status of several schemes from closed to open. Also, most of the open schemes represent large schemes, while closed schemes are often very small (see also Rama and McLeod 2001).

distinction between open and closed schemes did not matter for this analysis, since both their administration and risk pooling followed essentially similar rules under the Medical Schemes Act. The plan membership of each household is of much more interest, since it is the plan choice that might indicate adverse selection or risk selection.

Most of the 29 schemes, had only one plan option (39%), five schemes had two options (20%), two had three (12%), two had four (16%), and one had six options (12%). Table 4-2 compares the number of plan options and their percentage as a share of the different plan types.

Table 4-2: Comparison of number of plan options for all plan types

Number of plan options	Freq. all plan options	Percent of all plans	Cum.	Freq. open plan options	Percent of open plans	Cum.	Freq. closed plan options	Percent of closed plans	Cum.
One	19	38.78	38.78	1	4.35	4.35	18	69.23	69.23
Two	5	20.41	59.18	4	34.78	39.13	2	15.38	84.62
Three	2	12.24	71.43	2	26.09	65.22	0	0	84.62
Four	2	16.33	87.76	2	34.78	100	1	15.38	100
Five	0	0	87.76	0	0	100	0	0	100
Six	1	12.24	100	0	0	100	0	0	100
Total	**49**	**100**		**23**	**100**		**26**	**100**	

Within the nine open schemes there were 23 plan options. Two and three options were most frequent (35% and 26%, respectively). Most open medical schemes have two or more plan options among which members can choose. In comparison, among the nine closed schemes the vast majority (70%) had only one plan option. The list of the medical scheme plan options, together with their average number of observations over the four-year observation period is given in Appendix 4, table A4-2. From the previous observations that most open schemes are medium or large sized and most of the time have more than one plan option, it can be derived that medium and large schemes had most plan options. This can be confirmed with table A4-2 in Appendix 4.

B. Identifier variables: plan classification and group classification

For the analysis a household's plan membership was only considered in terms of the plan classification as a low-, average-, or high-risk plan.[77] The classification was done starting with the variable means averaged across all plans in the data set over the entire four-year

[77]Individual plan membership could not be otherwise considered, because the preferred model only allowed one clustering option. The model clustered observations from each household, obtained over the maximum

observation period, 1995-1998. Plans were then assessed to determine whether individual plan means for several variable values indicated a low-, average-, or high-risk accumulation, i.e. if plans comprised low-, average-, or high-risk pools. The plan classification into low-, average-, and high-risk plans was done applying the following process, the main components of which are also described in table 4-3.

Table 4-3: Group classification into low, high and average risk plans for the South African data

Classified plan type	Variables with expected plan means below overall mean range 95-98	Variables with expected plan means above overall mean range 95-98
Low-risk plan	age claims, contributions benefits (specifically basic benefits)	
High-risk plan		age claims, contributions benefits
Average-risk plan	no consistent pattern or within overall mean range	no consistent pattern or within overall mean range

First, the literature review, and in particular the risk-adjustment literature, concluded that it is possible to predict individuals' health expenditures (as a measure of risk for ill-health) using a few individual risk indicators. Specifically, easily obtainable measures of age and prior utilisation were identified as quite reliable measures (e.g. Newhouse 1996). This approach was applied here to classify plans into specific risk pools. The variable measures of age, contributions, claims, co-payments, and of several benefits were used for the classification of plans into low-, average-, and high-risk pools. For example, low-risk plans would be expected to show mostly below average values for age, contributions, claims, and benefits, in particular basic benefits as compared to the overall data means. High-risk plans were expected to display mostly above average values for age, contributions, claims, and benefits compared to the overall data means. Plans with average means within the overall data mean interval would be classified as average-risk plans.

Second, a list of variable means over the observation period 1995-1998 was compiled, averaged across all plans in the data set (and displayed in column one of table 4-4). These overall means were subsequently compared to individual plan means to determine whether the plan mean lay below or above the overall mean. However, rather than taking the single number of the overall variable mean as the cut-point for this

observation period of four years, in order to consider the panel character of the data. Another option - using 49 dummy variables for plan consideration - was dismissed for reasons of impracticability.

classification, a deviation, or mean interval, was defined around each overall variable mean (column three and four in table 4-4).

Table 4-4: Variable means and mean intervals for plan classification

Variables	Mean	Deviation	Mean interval
Socio-demographic			
age	42	4	38<42<46
Claims, contribution and benefits in Rand			
contributions	242	100	142<242<342
co-payment2	23	10	13<23<33
claims	229	80	149<229<309
Basic benefits			
GP	23	5	18<23<28
public hospital	2.0	1.5	0.5<2.0<2.5
Negotiable benefits			
specialist	36	15	21<36<51
private hospital	43	20	23<43<63
Discretionary benefits			
dental	17	5	12<17<22
optical	11	4	7<11<15
paramedical	3.3	2	1.3<3.3<5.3

Note: These are the means over all N=833,755 observations and n=325,957 households.

Third, the variable means for all individual plans were obtained and compiled in another table (see table A5-2 in Appendix 5). Next, all plan variable means that expressed a value above or below the overall mean were marked in this table. If, for example, a plan's variable mean for age lay beyond the upper mean interval border of 46, this would indicate for that specific plan that it accumulated older members, i.e. members with a high risk. If, on the other hand the plan mean age lay under the lower mean interval border of 38, this variable would then indicate that the plan pooled predominantly younger members, i.e. low risks. Yellow was used to mark plan means below the defined mean interval range. Pink was used to mark plan means above the defined mean interval range. Plan means that lay within the overall mean interval were left uncoloured and considered average risk pooling.

Finally, the identification of a *consistent pattern* over the variables in table 4-3 and 4-4 was focused on, either indicating an *accumulation of low risks or high risks.* If plans consistently showed the expected variable mean patterns that indicated an accumulation of low risks they were classified into low-risk plans. Claims and benefits as indicators of prior utilisation were identified by the risk adjustment literature as the best risk adjusters, while age and other socio-demographic variables still had some value as risk adjusters.

Thus the main variables of interest for the classification were, in order of importance: (1) total claims, (2) benefits, either supporting consistently a high or low risk pattern, and (3) age, as a supporting socio-demographic variable.

Out of the 49 plans 10 could be classified as low-risk plans. Similarly, if plan variable means were consistently above the overall mean range they most likely pooled high risks and could be classified as high-risk plans. Ten out of the 49 plans in the sample followed this high-risk pattern. The remaining 29 plans either displayed plan variable means within the overall mean range or other inconsistent patterns. These plans were classified as average-risk plans. The last column of table A5-2 in the Appendix 5 shows the classification of plans into low-, average-, and high-risk plans. Table A5-3 in Appendix 5 summarises the classification result of low- and high-risk plans - they display only plans and variable means for these two groups.

Last, but not least, a sensitivity analysis was performed by either increasing or decreasing the overall mean intervals, in order to confirm this classification. The sensitivity analysis was performed by increasing and decreasing the overall means intervals by 10% (see Appendix 5, table A5-1). Only a small number of mean classifications for all plans changed either by increasing or decreasing the means interval. Column 5 and column 7 in table A5-1 report these numbers. Increasing the mean interval by 10% altogether leads to only 12 mean classification changes for all 539 plan means. Decreasing the mean interval by 10% altogether leads to slightly more mean classification changes, 33 out of all 539 plan means. These low numbers suggest a very robust plan classification.

In Appendix 5 tables A5-4 and A5-5 show the plan classification into low- and high-risk plans with a 10% increased mean interval. Among low-risk and high-risk plans one mean classification changed for three plans. However, neither of these changes altered these plans overall classification as low- or high-risk plans. In Appendix 5 tables A5-6 and A5-7 show the plan classification into low- and high-risk plan with a 10% decreased mean interval. Among low-risk plans five mean classifications changed for three plans, but all of them supported their overall plan classification as low-risk plans. Among the high-risk plans five mean classifications changed for four plans. Here too, none of these mean classification changes altered their overall plan classification, on the contrary, some supported their classification as high-risk plans. Based on the sensitivity analysis it can be

concluded that the classification of the plans into low-, average-, and high-risk plans is robust.[78]

With this plan classification process it was possible to create a categorical variable, called risk (plan), that classified insurance plans into average risk carriers (and thus average-risk plans, taking the value zero) and plans with below (value 1) or above average (value 2) risks. About 70% of all households belonged to plans that had an average risk profile according to this classification. About a fifth belong to classified low-risk plans and only about 10% to classified high-risk plans.

The data set also contains households' entry and exit dates for their membership in *plans* and for their membership in *schemes*. This makes it possible to determine whether a household joined *medical schemes* during the observation period (i.e. belonged to a group of joiners) or was already insured in a medical scheme (and belonged to the control group) using a coded dummy variable. Similarly, it was possible to distinguish groups of leavers of *medical schemes* from the control group. Thus joiners and leavers *only* concern the *medical scheme* entry and exit. Switchers concern *only* the movement between different *plans* within a medical scheme. Altogether three dummy variables were created - joiners-control (jc), leavers-control (lc), and switchers-control (sc). These will be the independent variables in the three models estimated later. Table 4-5 summarises all four groups of households, the number of households in the sample used and the attributes of distinction between each

For example, for jc the value equals one if a household ever joined a *medical scheme* during the four-year observation period, zero if it belonged to the control group, and is marked differently otherwise (for instance if it belonged to the group of leavers or switchers). Thus joiners are *only* households that have a medical scheme entry date between 1994-1998.[79] Control group households are households with a medical scheme entry date before 1994 and no medical scheme exit date or plan switch date. The variable

[78]Originally this plan classification was done for each of the four years and these overall means. The yearly classification led to no changes in the final plan classification. Though some individual mean classifications for some plans did change, as happened with the sensitivity analysis. Because these tables are extremely space intensive, and do not provide a great deal of added information, and because the plan classification algorithm would have been further obscured by more table detail, the author decided to only use the overall means tables to describe the classification algorithm in this thesis. However, the yearly classification tables are available upon request from the author.

[79] Households were marked as joiners, leavers, or switchers for all available observations if they joined, left, or switched at any point in time during the observation period. This was done because the interest lays in the household's profile and behaviour over all available years of observation that either followed the joining and switching or preceded the switching and leaving. Due to this coding system the dummies are equivalent to time-invariant variables.

lc was similarly coded. Lc has the value one if the household ever left a *medical scheme* during the four-year observation period. Thus leavers are *only* households that have a medical scheme exit date between 1994-1998.

Table 4-5: Group classification into movers versus control group for the South African data

Group	Households	Attribute of distinction
Joiners (jc = 1)	73,941 (22%)	• Households joined medical schemes at any time during the observation period
Leavers (lc = 1)	97,729 (29%)	• Households left medical schemes at any time during the observation period
Switchers (sc = 1)	13,361 (3.7%)	• Households switched to another plan within a medical scheme during the observation period (about 90% of switchers entered a new plan)
Control (jc/lc/sc = 0)	164,620 (50%)	• Households were already insured and remained insured in original plan and medical scheme during the observation period

The coding for variable sc was slightly different. Switchers are households that entered or exited *plans* within medical schemes not *medical schemes* during the four-year observation period Thus switchers were always already members of a medical schemes (i.e. must have joined the medical scheme before 1994 and had such an entry date). Sc equals one if the household ever switched plans during the observation period. Thus switchers are *only* households that have a plan exit or entry date between 1994-1998.

It was pointed out that the variable sc was coded one if a household ever entered *or* exited a plan within a medical scheme. With the original data it was not possible to determine where switchers who entered a plan of a medical scheme came from or to where switchers who exited a medical scheme plan went. Most of the switcher observations (about 95%) described switchers who entered a plan. Because of the small number of switchers who exited plans, both types of switchers had to be pooled in light of the later statistical model analysis.[80] There will be no further distinction between switcher subgroups.

From table 4-5 the four data subgroups of interest can be derived, (1) the group of joiners, (2) the group of leavers, (3) the group of switchers and (4) the control group. These four data subgroups will be comparatively described in the next section, with consideration given to the panel data structure of the data.

[80] The discrepancy between the number of switchers that joined and switchers that left is explainable with filing negligence. For a plan manager it is far more important to open a file for a switcher that joined rather than to close the file of one that left. However, this has no effect on the later analysis.

4.2.3 Comparative data subgroup description considering the panel data structure

The joiners group counted 73,941 households, the leavers group 97,728, and the switchers group 13,361 (compared to 164,620 households in the control group and a total of 325,957 households) over the four years of observation. Table 4-6 demonstrates the panel character of the data, describing the data *T* pattern for the households in the joiners, leavers, switchers, and control groups, over the observation period *t*=1995, 1996, 1997, 1998. For example, the control group contains 164,620 households with uniquely identified observations over a maximum number of four household years.

Table 4-6: Data pattern over T for the control group versus mover groups

Control				Joiners			
Freq.	Perc.	Cum.	Pattern	Freq.	Perc.	Cum.	Pattern
78,205	47.51	47.51	1111	13,996	18.93	18.93	..11
13,339	8.10	55.61	1...	10,768	14.56	33.49	.111
10,802	6.56	62.17	11..	9,038	12.22	45.71	1111
10,599	6.44	68.61	111.	6,011	8.13	53.84	...1
10,368	6.30	74.91	11.1	5,767	7.80	61.64	11..
10,234	6.22	81.12	.111	5,588	7.56	69.20	.11.
7,462	4.53	85.66	1.11	4,885	6.61	75.81	1...
5,394	3.28	88.93	...1	4,809	6.5	82.31	..1.
5,060	3.07	92.01	..11	4,424	5.98	88.29	.1..
13,157	7.99	100	(other)	8,655	11.71	100	(other)
n=164,620	100.00		XXXX	n=73,941	100.00		XXXX
Leavers				Switchers			
23,360	23.9	23.9	11..	2,566	19.21	19.21	1111
22,786	23.32	47.22	1...	1,655	12.39	31.59	.111
17,707	18.12	65.34	111.	1,532	11.47	43.06	1...
9,192	9.41	74.74	.1..	1,358	10.16	53.22	...1
7,655	7.83	82.58	.11.	1,061	7.94	61.16	.1..
5,033	5.15	87.73	..1.	983	7.36	68.52	111.
3,920	4.01	91.74	1.1.	970	7.26	75.78	..11
2,983	3.05	94.79	1111	906	6.78	82.56	11..
1,443	1.48	96.27	..11	719	5.38	87.94	.11.
3,650	3.73	100	(other)	1,611	12.06	100	(other)
n=97,729	100.00		XXXX	n=13,361	100.00		XXXX

Note: In the pattern column the participation pattern is shown. A 1 in the pattern means one observation that year; a dot means no observation. The largest fraction of households, 47.5% of the control households, was observed for all four years, 1995, 1996, 1997 and 1998. The next largest fraction, 23.9% of leaver households, was observed in 1995 and 1996. At the bottom is the sum of the participation patterns; all data subgroups had households with all four years of observation.

It describes specifically the participation pattern of the cross-sectional time-series data. Of all households in the control group 47.5% have observations for all four time periods, while about 71% have at least three years of observations. Joiners have 12.2% of their households contributing with four-year observations, leavers only three percent, and

switchers 19.2%. However, all three have 27%, 21%, and 39% of their household observations, respectively, contributing at least three years.

The data of *n* households with recorded histories allows consideration of not only the household dimension within each cross-section, but also of households' intertemporal dynamics. For example, data with repeated observations is the only way to obtain information about individual patterns of change (Davis 2002). These intertemporal dynamics within panel data make the distinction between individual and average behaviour possible. The following comparative data subgroup description, whenever feasible, considers the panel structure of the data. This means that not only will the cross-sectional dimension of the data be regarded, but also that changes/trends over time will be considered. The subgroup households will be compared along different characteristics. These will be first, their socio-economic household characteristics, second their claim-contribution and benefits characteristics, and finally, the characteristics of their plan choices.

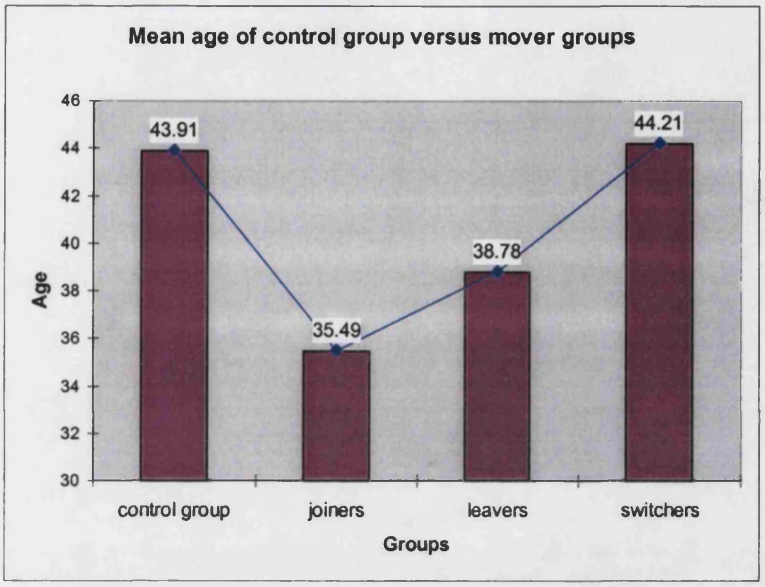
A. Socio-economic characteristics

The socio-economic characteristics were described by the variables of age, gender, race and income group for the three mover groups and the control group over the four year observation period.[81]

Figure 4-1 compares the mean age for the control group, joiners, leavers, and switchers. It is sufficient to work here with the overall subgroup means and it is not necessary to explore the panel structure for the age variable. This is because the age of the principal member will change naturally over the four-year observation period, but will not reveal any other enlightening trends. The overall mean age of the switchers is the highest of the four groups at 44.2 years, followed by the control group at 44.9. The joiners mean age is the lowest at 35.5 years. The overall mean age of leavers at 38.8 is above that of the joiners but still bellow the control group. Although the overall mean ages for the data subgroups are calculated from quite different observation numbers (ranging from 13,361 for the switchers to 501,158 for the control group), the standard deviation for each subgroup was around 12, the minima and maxima between zero and 98.

[81] Age was measured as a continuous value, while all other socio-economic variables were categorical.

Figure 4-1: Mean age comparison for control group versus mover groups



Even though overall two-thirds of all households are headed by a male principal member, female headed households account for nearly 40% of the joiners (compared to 28% in the control group, 32% in the leavers group and 26% in the switchers group). Taking the panel data structure into account, table 4-7 describes the one-way tabulation for the gender variable. The gender variable is one if the principal household member is female and is otherwise zero. The overall column refers to all observations – the household-years - while the between column refers to the households.

Although gender (like race) is a time-invariant variable, the within percent tabulation is not 100 and the total of the between column is higher than the *n* number of households.[82] This indicates that some households sometimes must have gen=0 and at other times gen=1, very unlikely for a time-invariant characteristic. However, this can be explained by the possibility that, for some households, the head of the household and thus principal insurance member changed during the observation period. This must have happened in rare cases (for instance, in 0.22% or 368 of all cases among the control group[83]) due to reasons like death or divorce. Subsequently another household member, such as the widow(er), took over the insurance plan as the new principal member. Generally, female headed households have higher proportions among the joiners and

[82] For time-invariant variables the within column should always show 100, since the variable value should stay constant over time (i.e. for all observations of one household). The within value is often interpreted as a stability measure.

leavers than the control group and lower proportions among the switchers than the control group.

Table 4-7: One-way table for gender variable for control group versus mover groups

Gender	Overall		Between		Within
Control	Freq.	Percent	Freq.	Percent	Percent
0 – male	364,146	72.66	119,554	72.62	99.77
1 – female	137,012	27.34	45,434	27.60	99.63
Total	**501,158**	**100.00**	**164,988**	**100.22**	**99.73**
			(n = 164,620)		
Joiners					
0 – male	99,010	60.93	45,185	61.11	99.93
1 – female	63,476	39.07	28,790	38.94	99.93
Total	**162,486**	**100.00**	**73,975**	**100.05**	**99.93**
			(n = 73,941)		
Leavers					
0 – male	121,866	66.31	65,901	67.43	99.97
1 – female	61, 915	33.69	31,851	32.59	99.95
Total	**183 ,781**	**100.00**	**97,752**	**100.02**	**99.96**
			(n = 97,729)		
Switchers					
0 – male	22,700	73.58	9,923	74.27	99.75
1 – female	8,152	26.42	3,467	25.95	99.56
Total	**30,852**	**100.00**	**13,390**	**100.22**	**99.70**
			(n = 13,361)		

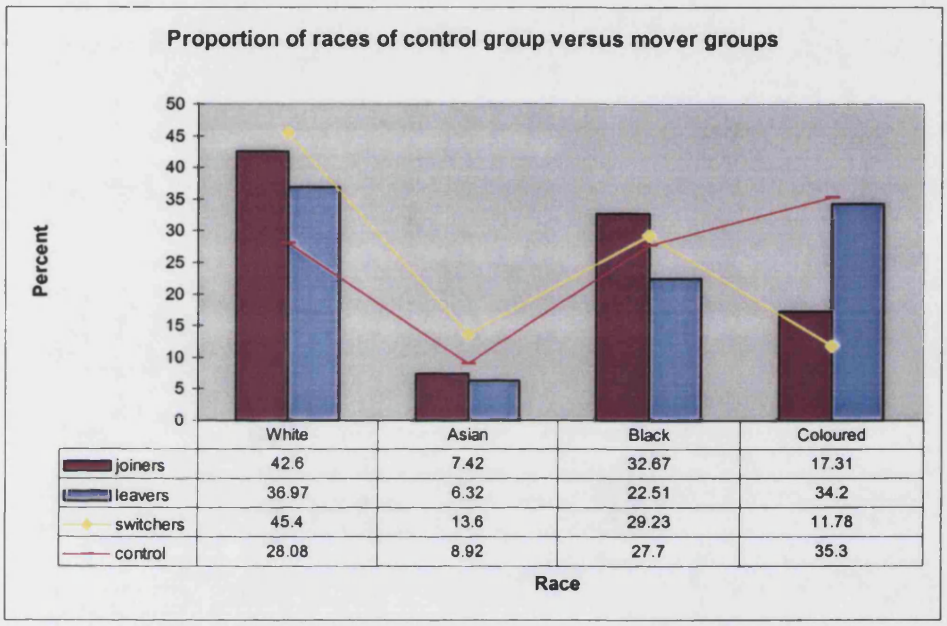
For the time-invariant variable race the proportion of each race within the four subgroups is displayed in figure 4-2.[84] Figure 4-2 shows that the race distribution over the four subgroups differs significantly. While the control group is dominated by Coloured households (39%), joiners and leavers have by far the highest White membership, 43% and 45% respectively. Whites also have the highest membership proportion among the leavers (37%).

It can be observed that first, Coloured headed households seem to enter and switch insurance less often, but also leave more often as compared to other groups with a similar control group representation (like Whites and Blacks). Thus their membership in the market is shrinking. Second, Black households account for the fastest growing racial group among the insured - Black joiners exceed the proporation of Black leavers and Black control group members. Third, the Asian membership group, although the smallest, turns out to be the most stable one with similar proportions among joiners, leavers and the

[83] The number 368 is the difference between the total number of observation in the between column and the number of *n* households in the data.

control group. However, Asian households switch plans comparatively frequently within the schemes of their membership. The proportion of White households, representing the largest race membership group, also seem the most volatile. Their proportion among the joiners is 15% higher and among the leavers nearly 10% higher than in the control group. White households also have the highest proportion of switchers between plans, being more prone to change their insurance coverage.

Figure 4-2: Race proportion comparison for control group versus mover groups



	White	Asian	Black	Coloured
joiners	42.6	7.42	32.67	17.31
leavers	36.97	6.32	22.51	34.2
switchers	45.4	13.6	29.23	11.78
control	28.08	8.92	27.7	35.3

Table 4-8 compares the one-way tables for the variable income group over all four data subgroups.[85] The income group variable is a typical time-variant variable and it is worthwhile to explore it regarding trends over time. In the control group about half of the households (46%) had an above average income at some time, while in the joiner group considerably fewer (38%) ever had an above average income. The overall and between values are quite similar for all four groups, although joiners seem to belong more often to the lower income groups. However, households in the control group belonged more often to the highest income group and also changed their income group most often.[86]

[84] Since race is a time-invariant variable, it will not change over time and thus, it is here sufficient to only assess the overall means over the four years of observation.
[85] Roughly three percent of all household-years in each of the groups have an average income, about 60% have a below average income, and 40% have an income above the average.
[86] The difference between the total number of households ever having either of the three income levels exceeds the *n* households for the control group most significantly, compared to all other groups. This is

Table 4-8: One-way table for income group variable for control group and movers

Income	Overall		Between		Within
Control	Freq.	Percent	Freq.	Percent	Percent
0 – average	12,500	2.49	7,926	4.81	50.29
1- below	286,875	57.24	105,433	64.05	88.51
2 – above	201,783	40.26	74,846	45.47	85.88
Total	**501,158**	**100.00**	**188,205**	**114.33**	**85.85**
			(n = 164,620)		
Joiners					
0 – average	6,405	3.94	4,297	5.81	61.49
1- below	100,113	61.61	47,695	64.50	93.14
2 – above	55,968	34.44	28,150	38.07	88.37
Total	**162,486**	**100.00**	**80,142**	**108.39**	**89.77**
			(n = 73,941)		
Leavers					
0 – average	5,399	2.94	3,963	4.06	69.07
1- below	107,471	58.48	60,355	61.76	92.46
2 – above	70,911	38.58	40,250	41.19	90.95
Total	**183,781**	**100.00**	**104,568**	**107.00**	**90.99**
			(n = 97,729)		
Switchers					
0 – average	1,216	3.94	744	5.57	65.27
1- below	16,362	53.03	7,872	58.92	91.77
2 – above	13,274	43.02	5,873	43.96	90.43
Total	**30,852**	**100.00**	**14,489**	**108.44**	**89.86**
			(n = 13,361)		

Note: Income group is one if a household belongs to a below average income group, two if it belongs to an above average income group and zero if it is in the average income group.

Table 4-9 gives an example of the transition probabilities that can be calculated in panel data for time-invariant variables, like income. The transition probability is the probability that the income group changes for a household in a following year given that there was a certain income value observation in the previous year. For example, if a household had an average income in 1994 it can change to a higher or lower income or stay the same in 1995. This transition probability is determined by counting transitions, i.e. a probability is calculating by counting how many people starting with any income level (average, low or high) in one year change their income level in the next year. The rows reflect the initial values and the columns the final values. The first row in table 4-9 shows the transition probability of an average income control group household for income changes during any year of observation. Each year 57.36% of average income control households remained in the average income group in the next year. Of the remaining 42.4%, for 2.24% the income fell below average, for 40.4% the income rose to above average.

supported by the values in the within percent columns, where the control group's comparably low levels indicate a lower stability of the lower income groups. Households with low income levels tend to improve income level more often if belonging to the control group than if belonging to any of the mover groups.

Table 4-9: Transition probability for income group variable for control group and mover groups

	Income			
Control	**0**	**1**	**2**	**Total**
0 - average	57.36	2.24	40.4	100
1- below	1.67	92.56	5.77	100
2 – above	0.41	5.09	94.5	100
Total	**2.48**	**55.63**	**41.89**	**100**
Joiners				
0 - average	62.87	3.58	33.54	100
1- below	2.36	93.98	3.66	100
2 – above	0.68	5.64	93.67	100
Total	**4.09**	**61.08**	**34.83**	**100**
Leavers				
0 - average	63.36	3	33.64	100
1- below	1.61	90.9	7.49	100
2 – above	0.62	4.02	95.36	100
Total	2.87	56.18	40.96	100
Switchers				
0 - average	75.56	2.89	21.54	100
1- below	2.21	93.41	4.38	100
2 – above	0.83	5.14	94.03	100
Total	**4.2**	**50.72**	**45.08**	**100**

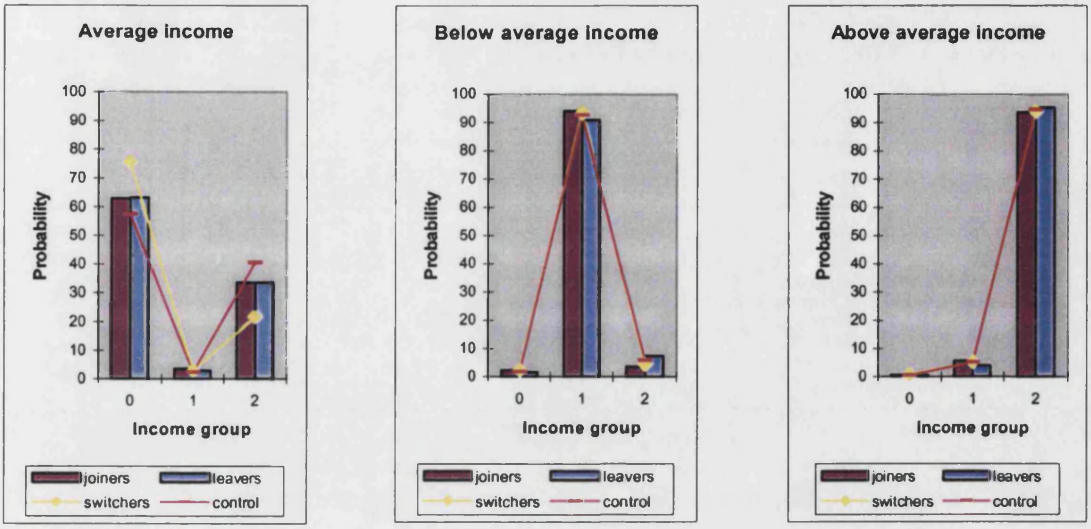
Note: The rows reflect the initial values and the columns the final values.

Figure 4-3 illustrates the transition probabilities of each income group for all subgroups. Each year some 60% of joining, leaving, and control group households and 75% of switching households with an average income remained at that income level in the next year. However, with 30-40% probability, joiners, leavers and control group members improve from the average income to the above average income level. Switchers have a comparably lower probability of only 22% for that transition.[87]

Those households starting at the below or above average income level generally have much lower transition probabilities. For example, below average income control group households had a 1.7% chance to become an average income household and a 5.8% chance to become an above average income household. Among households with an initial below average income, joiners are least likely to experience income growth as compared to all other groups. Leavers are most likely to achieve income growth and they also have the highest probability of remaining on the highest income level as compared to all other groups. However, these differences range only on a scale of one to three percent and do not reveal spectacular discrepancies between the groups.

[87] Thus the more switchers remain in the average income group, the less they progress towards an above average income group.

Figure 4-3: Transition probability comparison of income for control group versus mover groups



Note: Each of the three graphs displays the transition probabilities from one of the three different initial income values, while the columns in each graph reflect the final value. For example, the left figure displays the transition probabilities of the average income group (zero) to the below average income group (one) and above average income group (two). Thus, the probability values for the columns of income group zero are the probabilities of the subgroup household to remain in the average income group, if started out in it.

B. Claim-contribution and benefit characteristics

The claim-contribution and benefit characteristics of households from the three mover groups, joiners, leavers, and switchers and the control group, will be the variables of: claims, contributions, co-payments, and several medical benefits.[88] Although these are all time-variant variables, which should be described with respect to the panel structure, this is not feasible here due to the limited thesis volume. It is here of main interest to examine how their overall means vary over the four data subgroups.

Figure 4-4 compares the claim and contribution means of the control group with the three mover groups. The average claims (tac) in Rand are the highest for the groups of switchers and leavers, exceeding the claims of the control group. Only the joiners group has lower average claims as compared to the control group. The mean contribution (tco) levels of all mover groups exceed the average contribution of the control group. The contribution-claim ratio is above one for the joiners, which is higher than that of the control group. For leavers claims exceed the contributions (the contribution-claims ratio is below one) and are lower as compared to the control group.[89]

[88] All variables are continuous variables. The medical benefits are classified into essential emergency services like GP and public hospital benefits, non-essential negotiable services like specialists and private hospital benefits, and discretionary services like dental, optical and paramedical benefits. The variable values are measured in South African Rand per membership month and member.

[89] The switchers' contribution-claim ratio is one and nearly equals the ratio of the control group.

Figure 4-4: Mean claim and contribution comparison for control group versus mover groups

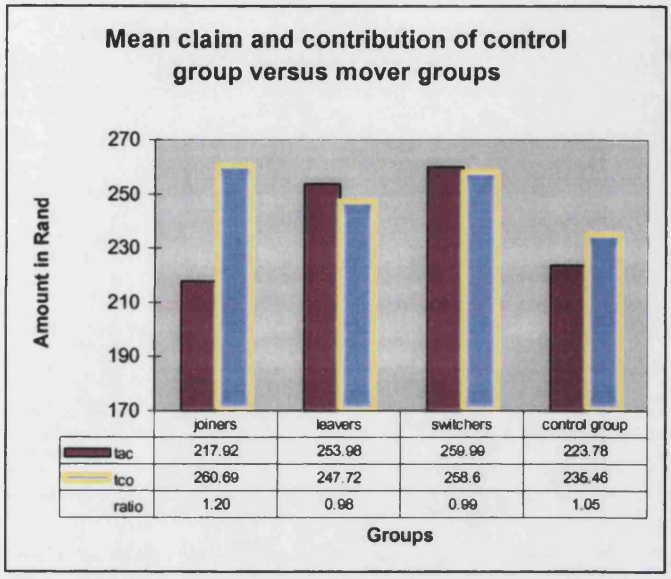
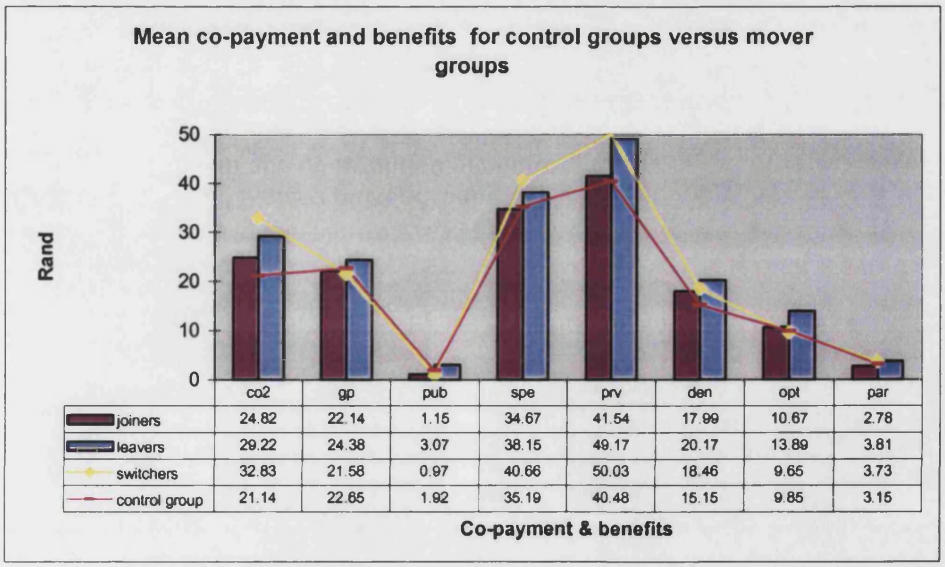


Figure 4-5 describes the comparison of the subgroups' mean values for the co-payment variable (co2), the essential benefit variables: GP (gp) and public hospital benefits (pub), the non-essential benefit variables: specialist (sp) and private hospital benefits (prv), and the discretionary benefit variables: dental (den), optical (opt), and paramedical benefits (par).

Figure 4-5: Mean co-payment and benefits comparison for control group versus mover groups



Leavers have higher average co-payments in comparison to the control group, but they also have higher essential benefits (being the highest among all mover groups). Switchers displayed higher average co-payments, lower essential benefits, but higher non-essential and discretionary benefits than the control group.

Taking the panel data structure into account, table 4-10 summarises as an example the descriptive statistic for dental benefits for all four subgroups. The overall mean for dental benefits is lowest for the control group (15 Rand) and highest for the leavers (20 Rand), spanning over a range from zero to a maximum of 1,162 Rand for the control group and 2,503 Rand for the leavers. The average dental benefit for each control group household varied between zero and 911 Rand, but was substantially more for leavers and joiners. The dental benefits 'within' (each household) varied for the control group between minus 374 and 739[90] and naturally some of those deviations from each household's average must be negative.

However, the deviation for the households of joiners and leavers is much higher - about a thousand Rand. Both, joiners and leavers did not only have higher dental benefits on average, they also deviated from their average much more than the control group households.

Table 4-10: Summary statistic for dental benefits of control group versus mover groups

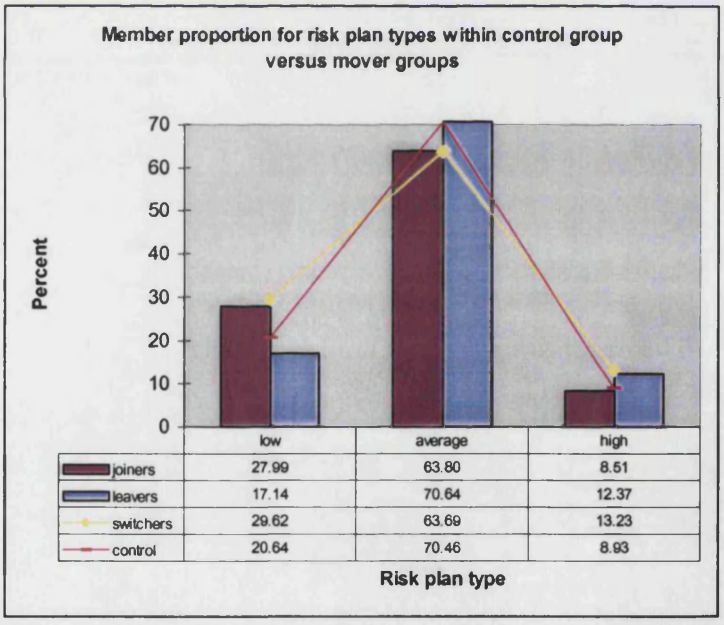
Dental benefits Control	Mean	Std. Dev.	Min	Max	Observations
overall	15.15	29.41	0	1,162	N = 501,158
between		22.58	0	911	n = 164,620
within		20.85	-374	754	T-bar = 3.0
Joiners					
overall	17.99	44.73	0	2,503	N = 162,486
between		39.40	0	2,118	n = 73,941
within		28.72	-1161	1,197	T-bar = 2.2
Leavers					
overall	20.17	55.10	0	2,503	N = 183,781
between		53.15	0	2,178	n = 97,729
within		31.36	-1182	1,223	T-bar = 1.9
Switchers					
overall	18.46	40.69	0	1,406	N = 30,852
between		37.64	0	932	n = 13,361
within		25.71	-685	721	T-bar = 2.3

C. Plan-choice characteristic

The plan-choice characteristic of households is described by the plan classification variable for a plan's risk type.[91] Although this variable is time-variant for the overall data, because households could switch between plans of different risk levels, for households within the different mover-groups the plan classification variable is in fact time-invariant. In the general data this variable could change if households switch between plans but remain within the medical scheme they originally belonged to.[92] However, the subgroups were constructed such that if a household moves at any time during the observation period all observations with that household identification (ID) were marked as a mover. That means each household that joined a plan is marked as a joiner for that specific plan over all years of its observation. Thus the risk plan variable is time-invariant for the data subgroups (as long as the household ID stays constant).[93]

Figure 4-6 displays the member proportions for each risk plan within the four types of households. The risk plan variable describes the risk level of the plans that the households belong to.

Figure 4-6: Member proportion for risk plan types within control group versus mover groups



Note: The risk variable has the value zero if a household had an observation for an average risk plan, one if the plan is classified as a low risk plan and two if it is of high risk.

[90] In the within definition the global average has to be added in again, here 754-15.15=738.85.
[91] The risk plan variable is categorical.
[92] However, it should be noted, that the risk plan variable would only capture some of these switchers, those that switch between plans of different risk levels.
[93] Since it is time-invariant for the subgroups, it is here sufficient to only regard the overall means over the four years of observation.

While in the control group about 70% of all households belonged to a plan with an average risk profile, switchers and joiners belonged to a similar plan only in about 60% of the cases. All groups had higher proportions of their observations in plans with a low-risk profile as compared to plans with high-risk profiles.[94] However, joiners and switchers more frequently belonged to a low-risk plan compared to the control group. Leavers belonged less often. And while leavers belonged more often to a high-risk plan before they exited, as compared to the control group, joiners were less likely to belong to a high-risk plan.

In summary, the comparative data subgroup description illustrated that the exploration of the data panel structure could reveal more exact and additional, valuable information regarding several issues of interest, particularly for time-variant variables. For example, income mobility and dental benefit patterns over time differ considerably between the data subgroups. This is because panel data not only describe cross-sectional characteristics of the observed households, but also changes over time. Thus panel data provide more data-points per household[95] and also allow the exploration of individual intertemporal dynamics. These advantages of panels, or longitudinal data[96], make it possible to analyse problems that can be hardly addressed by cross-sectional or time-series data alone (Hisao 1986). For example, testing of hypotheses regarding certain assumed household behaviours, like health care utilisation patterns in order to depict a household's risk type, becomes much more reliable with panel data, where a household's utilisation is observed over several time periods.

However, these special features of panel data can create new and difficult statistical problems, particularly in non-linear model building.[97] Specifically, they will

[94] Although this difference was highly in favour of low risk plans for the joiners (28% versus 9%) it was remarkably less pronounced for the leavers group (17% versus 12%) Switchers are less likely to belong to an average risk plan and tend to concentrate more in either low or high risk plans as compared to any other subgroup.

[95] These data also economise on households in many settings where it is desirable to observe the same household repeatedly rather than to observe different households at each point in time. Also data collection is arguably more reliable for studies in which the same households are followed repeatedly than in cross-sectional studies. There are other advantages for specific conditions and settings that involve experimental designs (see Davis 2002).

[96] The term longitudinal data is most often used by authors to describe repeated measurements data in which the repeated measurements factor is time. Longitudinal data, then, are a special case of repeated data. Others apply the term longitudinal data strictly to data that are collected over long time periods and uncontrolled conditions compared to repeated data for short time periods and experimental conditions. In this second case, repeated measurements data are a special case of longitudinal data (Davis 2002).

[97] Thus even though researchers agree that behaviour differs from one measured occasion to the next within each household, this intrahousehold variability of panel data is rarely explored. There are three aspects related to this fact. First, many statistical methods for panel data have only developed to a reasonable degree in the

require special statistical methods because each set of observations for each household tends to be intercorrelated[98] and data may be partially incomplete or unbalanced.[99] This means this thesis' empirical strategy needs to consider and accommodate these specific panel data requirements. The next section will describe the empirical strategy that considers the panel data structure of the South African data set.

4.3 Empirical strategy: a logit model for panel data

In order to pursue this thesis' second research aim of conducting an empirical analysis into adverse selection and risk selection the *empirical strategy* for the *analysis*, applying the group method to the case study data of South Africa's unregulated medical scheme market, needs to be specified. The purpose of this section is to describe the specific empirical strategy and statistical models for the South African panel data used in this thesis. While the last chapter depicted the general method of logistic regression as the best quantitative technique for the group method, the aim here is to identify the particular logit model structure that is best suited for the data at hand. First, this section will present an applied logistic regression model for correlated data - the marginal or population averaged model, which is the selected logistic regression model for the empirical analysis. The marginal model with its parameter estimation method and coefficient interpretation will be shortly introduced. Second, the model building strategy for this thesis' empirical analysis will be discussed, which will lead to the specifications of the marginal logit models applied. The aim is, in particular, to provide and derive variable expectations based on theory and previous research in order to address the thesis' research hypotheses. Finally, the model building process will be described. It will be explained how the final models were developed in order to establish that a credible model building process was followed, which resulted in the best models for this analysis.

---

last few years; second, their highly technical descriptions are perceived as inaccessible for applied researchers; and third, analysis software only became available recently (Moskowitz and Hershberger 2002).

[98] Repeated observations for the same household result in so-called clustered or correlated data.

[99] This concerns missing values in longitudinal data, although it is possible to distinguished between different reasons why they are missing. In general missing values arise whenever the response of a subject is missing for one or more time points and thus the number of repeated measures across subjects is not constant. Data are labelled incomplete if intended measurements are not taken, are lost, or otherwise unavailable (Diggle et al. 2002). Unbalanced data are a result of measurements that are made intentionally at an uncommon set of times on all subjects. Missing values resulting in incomplete and/or unbalanced data raise the same technical difficulties but relate to different conceptual issues.

4.3.1 Applied logistic regression with panel data: the marginal model

The data used for this thesis are panel data that contain household insurance data collected over four successive years. At the end of each year, each household's contribution, claims, benefits, and whether a household continued to be insured, or joined, left, or switched insurance during the current year were recorded. However, because households leave, join, and switch during the four-year observation period, not all households have observations for all four years, resulting in the problem of incomplete data. But since there is usually more than one (year) observation for each household, each household also represents a cluster of correlated observations. Thus as a second problem, the lack of independence in the observations arises because there is more than one year of measurements for each household.

The measurements of a household's contribution, claims, and benefits can change from observation to observation (i.e. they are time-variant) and they are called subject- or cluster-specific covariates. The household's principal member characteristics of age, gender, number of dependants, race, income level, and residential area are recorded when each household joins insurance, i.e. the first year of any record. Most of a household's principal member characteristics are constant (i.e. time-invariant) and they are called subject- or cluster-level covariates (Hosmer and Lemeshow 2000). However, since observations in the same cluster tend to exhibit intracluster correlation, standard analysis methods, like cross-sectional models, that ignore the clusters are inadequate. They lead to models with a poor fit and poorly estimated variances (Neuhaus et al. 1991).

Over recent years statistical research has developed several related types of extensions of the of generalised linear model (GLM) and quasilikelihood methods, specifically for the analysis of this type of correlated data (Davis 2002). Incomplete or missing data can also be accommodated with this model methodology.[100] The applied panel data literature discusses three extensions of GLMs for longitudinal data: (1) marginal models, (2) random effects models, and (3) transition (Markov) models. All three models have logistic formulations, however, each model is different in its underlying idea and the interpretation of the coefficients as well as their area of application.

This thesis' analytical aim is to characterise and contrast populations of households or subjects and the focus is therefore on the population averaged response. Marginal models assume that observations within each subject are correlated, but intraindividual

[100] However, the restriction here is that the missing data process is MCAR (missing completely at random) or MAR (missing at random).

correlation can be modelled separately from the regression of the response on explanatory variables. This model expresses the marginal expectation as a function of the explanatory variables. The intraindividual correlation is a known function of the marginal means and a vector of an unknown additional parameter. Thus in addition to modelling the effects of covariates on the marginal expectations, a model for the association among observations from each subject must be specified (Davis 2002). The interpretation of the estimated covariate coefficient β^* applies to a specific observed group with a shared attribute rather than individual subjects. This represents, for example, the effects of explanatory variables on that group of households' chances of joining or not.

Marginal models are also referred to as population averaged models (Zeger et al. 1988 and Neuhaus et al. 1991), because parameters only have interpretations in terms of the influence of covariates on the population averaged response. Thus marginal models are most effectively used in population studies like this, where the focus is the difference in the marginal or population average response between two groups. Or in other words, they are best fitted to settings where it is of interest to assess the effects of cluster-level or time-invariat covariates whose values cannot change. Since a marginal model is best suited for covariates that describe the effect of the covariate in broad groups of subjects, rather than individual subjects, it is, compared to the transition[101] or random effects model[102] most suited for this thesis' empirical investigation. However, in order to be able

[101] Transition models assume that observations within each subject are correlated because present observations are explicitly influenced by past outcomes. The model expresses the conditional mean of the present outcome as a function of covariates and past responses, giving past outcomes the role of additional predictors. The covariate effects are interpreted as being adjusted for the subject's response history (Davis 2002). The estimated covariate coefficient β^{**} of the transition model can be interpreted as the change per unit change in x in the log odds of joining, among households who did not join in the past (see Diggle et al. 2002). The effects of time-invariant covariates may be poorly estimated (Davis 2002). That makes transitional models most fitting to settings where it is of interest to assess the effects of time-variant covariates adjusted for the subjects' response histories. For example, this method would have been appropriate for this thesis if it would have been of primary interest to determine whether the chance of joining at a certain time depends, on whether or not households joined or left insurance coverage in the past. However, the focus of this thesis' empirical investigation will be not the effects of time-variant covariates adjusted for the households' response histories.

[102] Random effects models assume that observations within each subject are correlated because of their sharing of unmeasurable or unmeasured (omitted) variables and that this heterogeneity (or within-subject correlation) can be accounted for by including a subject-specific random term. Conditional on the values of the random effect, the repeated observations for each subject or the responses are assumed to be independent. The model also requires an assumption about the distribution of the random effects across subjects in the population. The covariate effects are interpreted in terms of their influence on both an individual's response and on the average response of the population (Davis 2002). Random effect models are also referred to as subject- and cluster-specific models (Zeger et al. 1988, Neuhaus et al. 1991) or latent variable models (Bartholomew 1987). They are best fitted to settings where it is of interest to assess the effects of cluster-specific or time-variant covariates whose values can change at the subject level, however, the interpretation of time-invariant factors is difficult (Hosmer and Lemeshow 2000). The estimated covariate coefficient β^{***} can be interpreted as the change per unit change in x in the log odds of joining for an individual household

to understand the interpretation of the estimated marginal logistic regression parameters it is useful to describe the statistical method employed by the marginal model.

It is assumed that the statistical method strives to describe an outcome y_{it}, for cluster (or subject) i at time t as a function of covariates $\mathbf{x}_{it}$ and parameters α_i specific to the ith cluster. Let y_{it} be a dichotomous random outcome variable and $\mathbf{x}_{it}$ a p x 1 vector of covariates at time t for subject i, where $t = 1,\ldots T$ and $i = 1,\ldots N$. Some of the covariates are constant within each subject (time-invariant/cluster-level variables), some may change from observation to observation (time-variant/cluster-specific variables), but there is heterogeneity across subjects. The marginal model then models the marginal expectation, $E(Y_{it})$, as a function of the explanatory variables. The marginal response is the average response over a subpopulation that shares a common value of x. The correlation between repeated observations is modelled separately from the marginal mean and variance of the response vector.

Marginal models have the following three assumptions (Davis 2002, Diggle et al. 2002): (1) the marginal expectation of the response, $E(Y_{it}) = \mu_{it}$ depends on explanatory variables and the relation can be described through a known link function $g(\mu_{it}) = \mathbf{x}_{it}'\beta^*$; (2) the marginal variance $\mathrm{Var}(y_{it})$ is related to the marginal expectation μ_{it} via a known variance function and a scale parameter; and (3) the correlation between observations is a known function of the marginal means and an additional parameter α^*.

The marginal model does not explicitly model subject-to-subject heterogeneity, the underlying logit model only specifies the marginal distribution of the averaged population response (modelled as a function of the covariates). Thus it models the marginal or population averaged distribution, rather than conditional distribution, given previous observations (Liang and Zeger 1986). The method is to average over the statistical distribution of the random effect term α. The statistical distribution of the random effects is not specified and the lack of distributional assumptions poses some problems when estimating the coefficients. The model coefficients apply to a regression model for a population with an averaged random effect α^* (and the interpretation applies to broad groups of subjects, not to a specific subject). A logistic marginal model is then given as:

(Diggle et al. 2002). However, the focus of this thesis' empirical investigation will be on groups of households with a shared attribute rather than on an individual household with a specific attribute.

$$\text{logit}\, P(y_{it} = 1 \mid x_{it}) = \alpha^* + \beta^* x_{it} \qquad \textit{(Equation 4-1: Marginal logit model)}$$

The estimated coefficient β^* of a population average logit model measures the change in the logit of the proportion with y=1 for a unit increase in x (Neuhaus et al. 1991). The form of the joint distribution of the repeated measurements is not fully specified and the estimation of β^* is approached with the semiparametric generalised estimating equation method (GEE) (Liang and Zeger 1986). The GEE approach is an extension of quasilikelihood to longitudinal data analysis and specifies the likelihood for marginal distributions and a working covariance matrix for the vector of repeated measurements from each subject.[103] However, this requires assumptions about the nature of the correlation. Most software packages offer four possible correlation assumptions: independent, auto-regressive, unstructured or exchangeable correlation (see also Hosmer and Lemeshow 2000, Davis 2002).

Under the independent model the correlation between the observations is zero, i.e. the correlation influence can be assumed to be small enough to fit a regression model for independent data. This correlation model leads only to consistent estimates of the parameter vector and covariance matrix given that the regression model is correctly specified. However, the correct variance of the parameters is often not supplied in standard software (Davis 2002). The assumption of no within correlation for this thesis' household observations seems quite unreasonable. The estimation of this model would then in fact be equivalent to the estimation of a standard logit/probit model. It would be only justifiable to use this correlation model if the correlation is low and the observations for each household are the same. In the model results section in chapter 5 the correlation matrix for each of the three models will indeed show that the correlation is not low. For these reasons the independent working correlation was not further considered.

The unstructured model, which assumes that the correlation between pairs of responses is different, would seem to be a good choice, because it is the most general and efficient model. However, this model generally requires the estimation of a large number of parameters that are of lesser importance here (Hosmer and Lemeshow 2000). Thus, although the estimation of unique pairwise correlation is very flexible, it is often not successful. Also, the occurrence of missing data complicates the estimation (Davis 2002). Each of this thesis' subgroups had a lot of data (several thousand households for an average of more than two years) and so estimating the full correlation matrix seemed

[103] See Liang and Zeger (1986) and Davis (2002) for good introductions to GEE.

feasible and was tried. However, models with an unstructured correlation did not converge. For these reasons the unstructured model was abandoned.

In the auto-regressive model with its time or order association within the observations, correlation decreases as the distance between the time points increases. Thus the correlation structure is defined as the usual correlation matrix for an AR(g) model. This is also called multiplicative correlation. These correlations decay very quickly as the spacing between observations increases. However, settings where there is an explicit time component are very specialised (Hosmer and Lemeshow 2000). When attempting to estimate the full correlation matrix for this thesis' models the resulting working correlation did not indicate a serial correlation of the residuals diminishing with increasing lags, which would imply an AR process. The correlation matrix did, however, look similar to an equal-correlation model, which corresponds to an exchangeable correlation structure. Thus the auto-regressive model was abandoned in favour of the exchangeable model.

The exchangeable correlation assumes a constant correlation between pairs of responses, meaning the correlation of observations within households is a constant. Under the assumption that observations covary equally within households, only a single parameter needs to be estimated, similar to a "random" effect. Also, this correlation model produces asymptotically valid standard errors even if the correlation structure is not truly exchangeable. In this thesis' analysis the interest lies solely in the estimation of the regression coefficients, but correlation between observations needs to be taken into account in order to obtain best estimates. For this case the exchangeable correlation model is most often recommended as the working correlation and was also chosen for this thesis' analysis.[104]

Liang and Zeger (1986) derive the so-called information sandwich estimator as an estimator of the covariance matrix. They show that the GEE yields consistent and asymptotically normal estimates of the regression coefficients and their variances even if the time dependence is misspecified (i.e. the assumed working covariance matrix). In addition, the loss of efficiency from an incorrect covariance matrix is inconsequential with large numbers of subjects as exist in this thesis' data (Davis 2002). The information

[104] The assumption of constant correlation between any two repeated measurements may not always be justified in a longitudinal study with a focus on changes over time. However, it is fitting if the responses in a household's cluster (are not naturally ordered and) were obtained from different family members living in that household, as is the case for this thesis' data (Davis 2002). For this thesis the focus was not on changes over time.

sandwich estimator, in most software packages referred to as the robust estimator, should always be used unless there is strong evidence for a correct specification of the working correlation structure from previous studies (Hosmer and Lemeshow 2000). This robust estimator is used in the thesis' analysis.

The estimated covariate coefficient β^* of the marginal model is interpreted as the change per unit change in x in the log odds of the proportion of households that join (or leave or switch).[105] The estimated odds ratios describe the risk of an event via proportions in the population and are much like odds ratios from other logistic regression models. Thus the estimated odds ratios of a marginal model are relatively simple to interpret since they compare odds computed from proportions of subjects in the population at different levels of the comparison covariate (holding all other variates fixed) (Hosmer and Lemeshow 2000).[106]

As an example, assume an odds ratio of 4 for a five-year age difference in people who leave insurance. Here the odds of leaving insurance coverage computed from the proportion of households whose principal members are five years older than some age reference level is four times higher than that based on the proportion of households whose principal members are at the reference age, holding all other variables constant. If age is linear in the logit the odds ratio holds for a five year difference at any age. The same interpretation is valid for the population average odds ratio of 2 for having a 10 Rand increase in basic benefits. The odds of being a leaver computed from the proportion of households who have 10 Rand more benefits than some reference level is 2% less than the odds of leaving based on the proportion of households who are at the reference, holding all other variables constant. The odds ratio for 10 Rand difference holds for any basic benefit level if basic benefits are linear in the logit. The examples showed that the marginal model is best used when the objective is to describe the effects of covariates in broad groups of subjects in the population. This broad interpretation comes at the cost of not using available information from repeated measures of the study subjects' covariates (Hosmer and Lemeshow 2000).

[105] Note that if all individuals with the same x have the same probability of joining/leaving/switching, the population frequency is the same as the individual's probability. However, when there is heterogeneity in the risk of joining/leaving/switching among subjects with a common x, the population frequency is the average of the individual risks (Diggle et al. 2002).

[106] However, a disadvantage is that this broad interpretation does not make use of information available in repeated measurements for each subject and that parameters depend on the degree of heterogeneity in the population.

The next section will describe the model building strategy and model specification for this thesis' marginal logistic regression models.

4.3.2 Model building strategy and process

The first step in a model building process is to select the method for specifying the models. The discussed data have a number of variables from which it is possible to select for inclusion in the logistic regression analysis. Here the confirmatory specification approach is applied, where the complete set of variables is prior-selected and the model is essentially used to confirm a hypothesis or theory.[107] This approach makes the model building process relatively easy at this stage, because the variable selection is determined by the literature, specifically literature that discusses methods of empirical investigations into selection.

The theoretical review of the selection processes and, in particular, the reviewed literature on empirical evidence for adverse selection and risk selection provided the underlying theory and hypotheses. First, the selection processes' theoretical concepts and the multiple plan choice literature, extended by the developed group approach, provided the theoretical hypothesis that adverse selection and risk selection are grouping phenomena that can be empirically identified within data subgroups. These are groups of subjects who join, leave, or switch insurance plans. Second, the risk adjustment literature provided the hypothesis that is it possible to determine a subject's risk using a risk profile comprised of different risk predictors (or variables that were called risk adjusters).

[107] A second, alternative, model building approach is the sequential search method, either in the form of the (1) stepwise estimation or the (2) forward addition and backward elimination. In the former the independent variable with the greatest contribution is added first. Independent variables are then selected for inclusion based on their incremental contribution compared to the variable(s) previously included (Hair et al. 1998). The stepwise approach allows addition or deletion at each stage. The later, forward addition is similar to the stepwise procedure, with the distinction that once a variable is added there is no reversion of that decision at a later stage. The backward elimination starts with the computation of the equation with all independent variables and then deletes the variable(s) that do not contribute significantly. Again, if a variable is once deleted, this approach does not include a reversion of that action at a later stage. Although this approach maximises the predictive ability of a model it can experience serious caveats in regard to the model interpretation. Since both of these approaches select variables one by one, it is highly unlikely that correlated variables will enter the model, which can lead to the misleading conclusion that not-included variable(s) are inconsequential. Excluded variables might be highly related to the dependant variable but also correlated to independent variables already in the model. A third model building approach is called the combinatorial approach, where the best fitting subset and combination of independent variables is selected as the model. Although this approach can identify the best overall equation with available software, it is not possible to address issues like multicollinearity, the effect of influential observations, and the general interpretability of the results with this method. However, these last two described general model building methods are not yet available for fitting correlated data models in commercially available software packages. This thesis is therefore left with the first approach: the confirmatory specification.

Empirical evidence for selection processes is concludes from unequal pooling of these risks in certain insurance plans.

This led to the thesis' main research hypothesis of a lack of risk pooling in South Africa's unregulated health insurance market due to adverse selection and risk selection. Within the case study setting of this thesis' this main research hypothesis was formulated into three subhypotheses. The key assumptions here were that it is possible to empirically identify adverse selection and risk selection because assumptions about the risk profiles of people within observed data subgroups can be made. If, for example, subjects with low-risk profiles join the medical schemes plans and particularly low-risk plans more, evidence for risk selection is likely.

The risk adjustment literature also provides these risk profile assumptions, and formulates specific variable expectations, in order to classify risks as low or high risks (and determine whether unequal risk pooling and empirical evidence for selection processes can be concluded). This means that the specific model building strategy used in this thesis will apply a confirmatory perspective with the purpose of verifying theoretical concepts and the developed group approach.

Thus the proposed logistic models have to explain membership in the defined mover groups using a multivariate risk profile and test whether or not these mover groups significantly differ in their risk profiles as compared to a control group. If the risk composition differs, evidence for either adverse selection or risk selection can be concluded. For the applied logistic regression analysis the required technical classification of observations into proposed groups that would provide the dependent variable was provided in this chapter's data section. The selection of the independent variables with their expectations (for the risk profile) of the proposed logistic models, were mainly based on hypotheses and results of previous empirical investigations from the risk adjustment and multiple plan choice literature, described in chapter two. Similar variables were described for the data at hand in the previous data section. This makes it possible to specify this thesis' analytical models. The following paragraphs will specify the models, summarise the variable expectations, and shortly describe the different steps of the model building process.

A. Model specification and variable expectations

Three discrete choice models were formulated, whereby the probabilities of different statuses of insurance coverage were hypothesised to be a function of household characteristics: $Logit[\text{insurance coverage status}] = \alpha + \beta' X + \varepsilon$

Where X is the vector of independent variables with the coefficients α and β that represent parameters to be estimated, and ε is a normally distributed random error term. The models were specified as marginal (or population averaged) logit models for panel data, which apply the semiparametric GEE in order to obtain parameter estimates. For the marginal models the within-group correlation structure for the panel was that of an equal-correlation model (exchangeable correlation). The information sandwich estimator or robust estimator of variance was used, producing valid standard errors, even if the within-group correlation structure was incorrectly hypothesised. The XTLOGIT, PA ROBUST command, with the clustering option over each household's ID, was used within the STATA™ statistical software system for computation (STATA 1999).[108] The three different models were specified with different dependant variables but the same set of independent variables.

Model A: Joiners versus control group $Logit[jc] = \alpha + \beta' X + \varepsilon$

Where the dependant variable jc indicates whether a household joined insurance coverage (i.e. plans) at any point during the observation period or belonged to the control group of observed households that never changed their insurance coverage over the four observed years (i.e. was already insured or remained insured in the same plan).

Model B: Leavers versus control group $Logit[lc] = \alpha + \beta' X + \varepsilon$

Where the dependant variable lc indicates whether a household left insurance coverage at any point during the observation period or belonged to the group of observed households that never changed their insurance coverage over the four years.

Model C: Switchers versus control group $Logit[sc] = \alpha + \beta' X + \varepsilon$

Where the dependant variable sc indicates whether a household switched insurance plan coverage at any point during the observation period or belonged to the group of observed households that never changed their insurance coverage over the four years.

Table 4-11 describes the vector X of independent variables. These independent variables are also referred to as the main effects.

[108] The STATA™ command XTGEE, FAMILY(BIN) LINK(LOGIT) CORR(EXCH) ROBUST estimates the same population averaged model.

Table 4-11: Description of vector of independent variables

Variable Name	Variable Description
age	Age of principal member
gen	Gender of principal member – female (1), male (0)
dep	Number of dependants of principal member (being household members)
inc	Income group of principal member – below average (1), above average (2), average (0)
race	Race of principal member – Asian (1), Black (2), Coloured (3), White (0)
urb	Residential area – urban (1), rural (0)
tac	Claims per household member and month in Rand
tco	Contributions per household member and month in Rand
co2	Co-payments (=legally billable amount minus benefits) per household member and month in Rand
gp	GP benefits per household member and month in Rand
pub	Public hospital benefits per household member and month in Rand
spe	Specialist benefits per household member and month in Rand
prv	Private hospital benefits per household member and month in Rand
den	Dental benefits per household member and month in Rand
opt	Optical benefits per household member and month in Rand
par	Paramedical benefits per household member and month in Rand
risk	Risk type of plan – low risk (1), high risk (2), average risk (0),

The following paragraphs will describe the variable expectations under adverse selection and risk selection. These expectations are mainly derived from previous research by this author and from the risk-adjustment literature. They address the thesis' research hypothesis, that unregulated health insurance markets encounter an unstable market-wide risk pool and imbalanced risk pools within the market (i.e. between insurers' plan options). Most of these variables are assumed to deliver some evidence for unequal risk pooling in the South African unregulated medical scheme market.

First, socio-economic variables like age, gender, race, and residential area were concluded to be of some use as risk indicators in the risk-adjustment literature (Newhouse 1996). For example, aside from the fact that older people might leave more and younger people might join more for natural reasons, it can be still expected that the risk for ill-health rises with age and thus insurance coverage is then either harder to obtain or harder to keep (Hansl and Soderlund 1999, Soderlund and Hansl 1999). From this it could be hypothesised that it is likely that older people adversely select themselves into insurance coverage and are also more likely to be unfavourably risk selected (i.e. dumped) by insurers. Younger people are typically targeted by insurers' efforts to attract low risks. Thus under adverse selection it is assumed that joiners are older and leavers are younger than control group subjects. However, if leavers are younger adverse selection might have been induced through the favourable risk selection efforts of external private, commercial insurers. In this case it will be necessary to assess whether or not there is information

regarding aggressive recruiting by external competitors. Adverse selection might also be operating if switchers are older than control group subjects, i.e. high risks chose to switch (to better coverage and thus high-risk plans). Comparably, under risk selection it is expected that joiners are younger and leavers are older as compared to the control group. However, this observation might be also influenced by natural life-cycle decisions and events, i.e. people join insurance coverage when they are young, starting their professional life and exit insurance coverage when old due to death. In the case of favourable risk selection switchers are expected to be younger than control group subjects (and switch to low-risk plans). Also, unfavourable risk selection is suspected if switchers are older (and switch to high-risk plans).

For gender and dependant number similar expectations can be formulated. Unfavourable risk discrimination against female insurance applicants and existing members is likely, because higher utilisation levels in reproductive years and in old age can be expected. The same reasons might induce females to seek insurance, which can result in adverse selection. Under adverse selection it is expected that females are more prominent among joiners and males are more prominent among the leavers as compared to the control group.[109] Subjects with a high number of dependants are of higher risk to induce costs because more people are covered under one insurance contract. They are expected to have more difficulties obtaining and keeping insurance. Insurers are expected to seek low-risk singles. Under risk selection it is expected that joiners have fewer dependants and leavers have more as compared to the control group. Under adverse selection it is assumed that joiners have more dependants and leavers fewer in comparison to the control group. Again, fewer dependants in the leavers groups might also indicate favourable risk selection by external private competitors.

Income[110], race,[111] and urban residency[112] are not expected to be the basis for risk discrimination in insurance coverage. They are, nevertheless, important control variables and should be included in the statistical analysis for that reason.

[109] However, gender biases also occur in employment and compensation practices such that men might have generally higher level of insurance coverage, whereas women are expected to receive coverage via their spouses (Hansl and Soderlund 2000).
[110] Income might function as an affordability constraint for subjects' ability to obtain and to keep insurance. Empirical evidence suggests that the demand for health care increases with increasing income, and thus one would predict increasing levels of insurance coverage with increasing income (Soderlund and Hansl 1999). However, for low-income earners it would be rational to seek subsidised employment-based insurance because illness will be more likely to have ruinous effects on their households' limited disposable income.
[111] Historic inequities, prejudices on the part of employers, and lower demand for western medical care among Non-whites could lead to higher levels of insurance coverage amongst whites, other factors being equal (Soderlund and Hansl 2000).

Second, variables that are some function of prior utilisation are graded by the risk-adjustment literature as potentially the best risk indicators (Newhouse 1996). Within the context of this analysis, variables that might reflect prior utilisation include: claims, contributions, co-payments, and different benefits. Medical benefits will be distinguished into essential emergency benefits, like GP and public hospital benefits, non-essential negotiable benefits like specialists and private hospital benefits, and discretionary benefits like dental, optical, and paramedical benefits.

For example, claims indicate whether or not a subject substantially utilised health services and is thus a low or high risk. Insurers would always prefer subjects with prospective low risk, since they keep costs low. Insurers are also eager to get rid of subjects with high claims. Thus under adverse selection joiners are expected to have higher claims and leavers lower claims as compared to the control group. However, low claims in the leavers group might also indicate favourable risk selection by external private competitors. Switchers are expected to have higher claims as compared to control group subjects under adverse selection of high risks (into high-risk plans). Under risk selection, however, it is expected that joiners will have lower claims and leavers higher claims than the control group. Switchers are expected to have lower claims as compared to the control group in the case of favourable risk selection (into low-risk plans). Also, unfavourable risk selection is suspected if switchers have comparably high claims (and switch to high-risk plans).

Contributions and co-payment are, on their own, less able to explain or predict subjects' risk and thus less useful as factors for risk discrimination. Contributions, for example, could be only used by South African medical schemes to purposefully discriminate between risks since the deregulation policy was implemented in the early to mid 1990s. In cases of unfavourable risk selection one would expect leavers and switchers to have higher contributions than the control group (because insurers can charge high risks higher contributions). However, for new insurance applicants the expectations are less straightforward, since insurers cannot really predict their future claim levels. It might be expected that the economic incentive of profit-maximisation brings them to charge relatively high premiums for new members, leading to the observation of generally high

[112] Urban residency might provide an indication of subjects' access to insurance coverage and health care services. Since urban dwellers have better access to both one would predict increasing levels of insurance coverage with urban residency. Therefore, insurers would rather select rural dwellers into insurance since their likelihood of cost occurrence is lower (Newhouse 1996). However, rural households are hard to access

contribution levels for joiners (even if among these are favourably risk selected low risks) compared to the control group.

However, a better indicator is the contribution-claim ratio. If, for example, the ratio is above one, i.e. claims exceed the subjects' contributions, insurers would classify this member as a high risk. Similarly, if the contribution-claim ratio is below one the subject is a low risk in the eyes of the insurer. Thus, under adverse selection, joiners are expected to have lower and leavers higher contribution-claim ratios than the control group. However, for the model estimations in the analysis it was decided to only include one of the two, either the contribution variable or the contribution-claim ratio. The models will consider the contribution variable only, while the average contribution-claim ratio for each mover group was presented in this chapter's data section.

Co-payment is in some way a function of health care utilisation, because in most cases of ill-health health services are used, costs are induced, and shared costs may arise. The problem is that co-payment rates vary for different types of services, but the co-payment variable itself cannot specify on what type of medical service it was spent. Also, the co-payment variable gives no indication about the total underlying claim or benefit level.

For example, a high risk would spend co-payments on essential services that usually have low co-payment rates, while a low risk would probably spend it on discretionary services, for which co-payments can be substantial. Lets say a low risk gets a massage that costs 20 Rand and the co-payment rate for physiotherapy is 50%. This would cost both the insurer and the insured 10 Rand. Similarly, lets assume a high risk gets a prescription for a chronic condition filled by a GP for 100 Rand with a co-payment rate of 10%. This would cost the insurer 90 Rand and the insured 10 Rand. Thus each risk type might end up with the similar co-payment but very different total claims and benefits. Also, each risk type would have spent their co-payment on different service types.

This shows that the co-payment variable alone cannot be used to distinguish who is a high risk and who is a low risk without also considering other claim and benefit variables. First, high claims with any level of co-payments indicate a high risk, while low claims always will indicate a low risk, regardless of the co-payment level. Second, high essential benefits with any level of co-payments indicate a high risk, while low essential benefits always will indicate a low risk, regardless of the co-payment level. The expected

for insurance underwriters and these households also have little incentive to take out insurance if they cannot easily access health care facilities.

observations regarding adverse selection and risk selection have to be based on observations for these claim and benefit variables. Nevertheless, they were included as control variables in order to be assessed together with the claim and benefit variables.

The remaining seven variables of this group are different benefits variables that would allow a more qualified assessment of the kind of health care services used and thus what risk type a household belongs to. Benefits for essential emergency services (i.e. GP and public hospital benefits)[113] are expected to be good risk indicators for high risks. Benefits for negotiable non-emergency services (i.e. specialist and private hospital benefits) and benefits for discretionary services (i.e. dental, optical and paramedical benefits) are expected to be of lesser significance for high risks. In fact, negotiable and discretionary services are expected to be more prominent among low risks, because high standard services function as luxury goods. They are offered preferentially to low risks in order to make their insurance costs justifiable and thereby keep them insured. Otherwise the expectations for negotiable and discretionary benefits are identical to the essential services. Subjects with low benefit patterns are of low risk and are more attractive to insurers, because of their low costs.

Finally, the plan classification variable that classifies plans with regard to their risk pool is expected to be a very good indicator for risk discrimination. It is expected that subjects either self-select themselves or are selected by the insurer into specific plan risk pools. For example, under adverse selection it is assumed that joiners are more likely to be high risks and belong more often to high-risk plans, leavers are more likely low risks who more often belong to low-risk plans as compared to control group subjects. Under risk selection it is expected that joiners are more likely low risks and belong more often to low-risk plans, and leavers are more likely to be high risks and belong more often to high-risk plans as compared to the control group.[114]

The following table 4-12 and table 4-13 summarise the expected observations under adverse selection (AS) or risk selection (RS) using the developed typology for adverse selection and risk selection from chapter three.

[113] Subjects with high essential benefits are costly for the insurer and are more likely to be dumped, while insurers are interested in recruiting people with prospective low essential benefits. The expectations for the essential benefits variables are similar to those of the claim variable. However, public hospital services is deemed the better indicator of risk, because in many instances GPs function as natural gatekeepers for non-essential and discretionary services that are more interesting for high risks.

[114] If switchers are of high risk and switch more often to high coverage (i.e. high-risk plans) as compared to control group subjects adverse selection might be operating. switchers are expected to be of low risk and belong more often to low-risk plans in the case of favourable risk selection. If switchers are of high risk and switch to high-risk plans unfavourable risk selection is suspected.

Table 4-12: Expected empirical observations under adverse selection

Adverse Selection	**Type 1** **Choice of insurance coverage by high risks**	**Type 2** **Choice of certain plans of insurance coverage by high risks**	**Type 3** **Choice of discontinued insurance coverage by low risks**
MANIFESTATION	Entrance of High Risks	**A.** Entrance of High Risks into High-Risk Plans **B.** Switch of High Risks to High-Risk Plans	**A.** Exit of Low Risks **B.** Exit of Low Risks from High-Risk Plans
MECHANISMS	ADVERSE SELECTION OF HIGH RISKS	ADVERSE SELECTION OF HIGH RISKS	ADVERSE SELECTION OF LOW RISKS
AFFECTED GROUPS	JOINERS	**A.** JOINERS **B.** SWITCHERS	**A.** LEAVERS **B.** LEAVERS
EXPECTED EMPIRICAL OBSERVATIONS	On average Joiners: 1. Are older, female, non-single 2. Have higher claims, lower contribution-claim ratios 3. Have higher essential and non-essential benefits	A. On average Joiners: 1. Are older, female, non-single 2. Have higher claims, lower contribution-claim ratios 3. Have higher essential and non-essential benefits Join high-risk plans B. On average Switchers: 1. Are older, female, non-single 2. Have higher claims, lower contribution-claim ratios 3. Have higher essential and non-essential benefits Switch to high-risk plans	A. On average Leavers: 1. Are younger, male, single 2. Have lower claims, higher contribution-claim ratios 3. Have lower essential, non-essential benefits and higher discretionary benefits B. On average Leavers: 1. Are younger, male, single 2. Have lower claims, higher contribution-claim ratios 4. Have lower essential, non-essential service utilisation and higher discretionary benefits Leave from high-risk plans

For example, adverse selection type 2A can be suspected (table 4-12, column two) if it is observed that subjects in the joiner group are older, female, non-single, have higher claims, co-payments, essential and non-essential benefits, and belong more often to high-risk plans as compared to the control group. As described in table 4-13 (column two) risk selection of type 2A can be suspected, if it is observed that joiners are more likely to be younger, male, single, have lower claims, essential and non-essential benefits, but higher discretionary benefits, and belong more often to low-risk plans as compared to the control group.

Table 4-13: Expected empirical observations under risk selection

Risk Selection	Type 1 **Discourage high-risk insurance applicants**	Type 2 **Attract low risks to certain plans of insurance coverage**	Type 3 **Discourage high risks to continue insurance coverage or switch them to high-risk plans**
MANIFESTATION	Restricted Entrance of High Risks	**A.** Entrance of Low Risks into Low-Risk Plans **B.** Switch of Low Risks to Low-Risk Plans	**A.** Exit of High Risks (from High-Risk Plans **B.** Switch of High Risks to High-Risk Plans
MECHANISMS	UNFAVOURABLE RISK SELECTION	FAVOURABLE RISK SELECTION/CREAM-SKIMMING	A. DUMPING B. UNFAVOURABLE RISK SELECTION
AFFECTED GROUPS	JOINERS	**A.** JOINERS **B.** SWITCHERS	**A.** LEAVERS **B.** SWITCHERS
EXPECTED EMPIRICAL OBSERVATIONS	On average Joiners: 1. Are younger, male, single 2. Have lower claims, higher contribution-claim ratios 3. Have lower essential, non-essential and higher discretionary benefits	A On average Joiners: 1. Are younger, male, single 2. Have lower claims, higher contribution-claim ratios 3. Have lower essential, non-essential and higher discretionary benefits Join low-risk plans B. On average Switchers: 1. Are younger, male, single 2. Have lower claims, higher contribution-claim ratios 3. Have lower essential, non-essential and higher discretionary benefits Switch to low-risk plans	A. On average Leavers: 1. Are older, female, non-single 2. Have higher claims, contributions, lower contribution-claim ratios 3. Have higher essential and non-essential benefits Leave high-risk plans B. On average Switchers: 1. Are older, female, non-single 2. Have higher claims, contributions, lower contribution-claim ratios 3. Have higher essential and higher non-essential benefits Switch to high-risk plans

B. Model building process

A model building process was selected that folloes four stages:

(1) Variable inclusion verification with univariate model assessment

(2) Model variable check for linearity in the logit

(3) Model check for interactions

(4) Assessment of model fit

This four-stage model building process is established in the literature for applied logistic regression (Hosmer and Lemeshow 2000) and was employed for each model described in the preceding section. The model building steps (1) to (3) are mapped out in detail in Appendix 6.

First, a univariate model assessment with each selected variable was performed in order to verify the variable inclusion. This first step is particularly important because a

confirmatory model specification approach was used for the variable selection. Therefore, it is necessary to assure that the selected set of variables achieves the maximum prediction while maintaining an efficient model. The use of the Wald tests for marginal models with correlated data is advised here (Hosmer and Lemeshow 2000). The Wald test found some association with the outcome for most variables in all three models, which justified the initial variable selection for the multivariate analysis and resulted in the so-called main effects model.

Second, the model variables were checked for linearity in the logit, which is a key assumption for any logistic regression analysis. Upon completion of this model building step, the initial variable selection for each model remained the same, confirming the main effects model.

Third, the models were checked for interactions, another important assumption for logistic regression analysis. Any interaction between two variables implies non-constant values for one variable over levels of the other. For example, an interaction term between gender and income entails that the slope coefficient for income is different for females and males. During this model building step, several interaction terms needed to be added to each main effects model, resulting in the so-called preliminary final models. Each of the three preliminary final models, Models A, B and C, had 17 similar variables and 15 different interaction terms. These models were preliminary final models because they were not yet checked for their model fit.

The final step of the model building process, the model fit assessment, will be presented in the next chapter after the discussion of the model results. The next section will summarise this chapter.

4.4 Summary

This chapter presented the data and empirical strategy used in order to pursue this thesis' second research aim of conducting an empirical analysis into adverse selection and risk selection applying the group method. The data used for this empirical analysis describe a panel data sample of an unregulated health insurance market in a middle-income country, South Africa's medical schemes between 1995-1998.

The first part of this chapter discussed the data and data sample derivation. First, a representative data sample for South Africa's unregulated medical schemes between 1995–1998 was derived. It contains of n =325,957 households belonging to 29 schemes with 49 plan options. The data provide detailed information on a series of households' socio-

economic characteristics, contribution, claim and benefit characteristics, and the dates when a chosen scheme and plan coverage started and ended. Two crucial variables for the later analysis were created, the plan classification variable that classified plans into different risk pools and the data subgroup (or mover group) identifier variables. Based on the subgroup identifier variables the data subgroups could be defined and described. The comparative data subgroup description took the panel data structure into account, which provided additional valuable information on households' intertemporal dynamics, like income mobility or claim and benefit patterns. While panel data make hypothesis testing for household behaviour more reliable, each set of household observations tends to be intercorrelated and sometimes incomplete, which requires special statistical methods for the analysis.

The second part of this chapter presented the empirical strategy for this thesis' panel data analysis using logistic regression. First, among the applied statistical models for this type of correlated data, which also have logistic formulations, the marginal or population averaged model was identified as the best suited. The empirical method using the marginal logit model specifies the likelihood for marginal distributions and a working covariance matrix for the vector of repeated measurements from each subject. The parameter estimation is approached with the GEE method. Software packages that can fit marginal models are widely available and thus easily accessible. The presented model building strategy followed a confirmatory approach and led to the model specifications. Three models were specified to bring about evidence for adverse selection and risk selection in the unregulated South African medical scheme market. Model A discriminates between joiners and the control group, Model B discriminates between leavers and the control group, and Model C discriminates between switchers and the control group. The selection of the independent variables (for the risk profile) of the proposed logistic models and their formulated expectations were based on hypotheses from previous empirical risk adjustment studies, the multiple plan choice literature. A four-step model-building process was applied in order to ensure that the specified models fulfilled the assumptions of logistic regression, i.e. that they are efficient and predict well. The results were three preliminary final models, Models A, B, and C, each with 17 similar variables (or main effects) and 15 different interaction terms. These models were deemed preliminary final models because they still need to be checked for their model fit.

The results of the models developed for this thesis' empirical investigation will be presented and interpreted in the next chapter.

Chapter 5

Analysis results & interpretation

5.1 Introduction

The purpose of this chapter is to fulfil this thesis' research objective by presenting the results of the empirical investigation and by assessing the research hypotheses. The empirical analysis aimed to either verify or falsify the main research hypothesis that South Africa's unregulated health insurance markets experience a lack of risk pooling. This main research hypothesis was formulated into several subhypotheses. First, it is the aim of this chapter to present the results of the statistical models that were specified in the previous chapter. Second, the results for the model fit assessment will be presented, comprising several standard diagnostic statistics. Third, the results of the three models will be interpreted in regard to their evidence for adverse selection and/or risk selection. Finally, it will be the aim of this chapter to summarise and review the empirical analysis results, particularly in regard to the previously formulated research hypotheses.

After this introduction, part two of the chapter will present the model estimation results. The goal here is to describe and explain the results for the logit models, which are based on estimated probabilities and odds ratios. Each model will enable the assessment of risk profile differences between a control group of constantly insured individuals and, in Model A, medical scheme plan joiners, in Model B, leavers, and in Model C, switchers.

The third part of this chapter will conclude the results for the statistical analysis with the results for the model fit assessment. However, since diagnostic statistics for the overall fit of logistic regression models for correlated data have not yet been developed the model fit assessment will be approximated with standard methods developed for normal uncorrelated data. The aim here is to apply several standard model fit assessment methods to the three logit models by assuming that the observations are not correlated, as

opposed to not performing any model checking at all. This will, however, limit the interpretation of the model fit assessment results, since they will not represent the model fit results of the actual marginal models, but their logistic proxy models. The applied model fit assessment methods range from summary measures of goodness-of-fit to logistic regression diagnostics.

The fourth part will interpret the model results and assess evidence for adverse selection and risk selection in South Africa's unregulated medical scheme market between 1995-1998. It will begin with a summary and synthesis of the empirical analysis results within the group method framework. It will be shown that evidence for adverse selection is very limited and the typical adverse selection cycle cannot be confirmed. However, there is overwhelming evidence for risk selection, in the form of discouraging the entry of high risks, cream-skimming of low risks and dumping of high risks. The results interpretation will then be discussed for all models and variables, including socio-economic variables, claim-contribution & benefit variables, and the plan classification variable. It will be demonstrated that the results of all three models can be interpreted as evidence for either adverse selection or risk selection.

Finally, the fifth part will review the model estimation results in the context of this thesis' research hypotheses. This assessment will verify this thesis' main research hypothesis, that South Africa's unregulated health insurance market encountered a lack of risk pooling.

5.2 Results of model estimation

The purpose of this section is to present the results of this thesis' empirical investigation into evidence for adverse selection and risk selection with data describing the unregulated medical scheme market in South Africa. First, this section will present the estimation results of Model A, which assesses differences in the risk profiles of medical scheme plan joiners and the control group of the constantly insured in medical schemes. Second, the results of Model B will be described, which aimed to discriminate between leavers from the medical scheme plans and the control group. Finally, the results of the third logistic regression model are presented. This final model assesses differences in risk profile characteristics of those who switch between medical scheme plans and the control group.

5.2.1 Model A results: joiners-control group

Model A tested for differences between the risk profiles of joiners (households that join medical scheme plans during any time during the observation period) and the control group (households constantly insured in medical scheme plans over the whole observation period). The dependant variable for Model A is jc, with jc=1 for households that joined at any point during the observation period and zero otherwise. Table 5-1 on the next page displays Model A's coefficients, odds ratios, semi-robust standard errors, z-scores, two-tailed *p*-values, and 95% confidence intervals for the population average model, with exchangeable correlation. The 663,644 observations that entered Model A were clustered over 238,561 households. The minimum of observations per household is 1 (year) and the maximum is 4 years, yielding an average of 2.8 observations per household. Model A's covariates are all at least significant at the five percent level as shown in table 5-1. The following estimated within household working correlation matrix R can be reported:

	c1	c2	c3	c4
r1	1.0000			
r2	0.8642	1.0000		
r3	0.8642	0.8642	1.0000	
r4	0.8642	0.8642	0.8642	1.0000

The interpretation of the estimated odds ratios from this population averaged model is straightforward, since it compares odds computed from proportions of subjects in the population at the different comparison covariant levels holding all other covariates fixed. For example the estimated population average odds ratio for a female headed household (gen) is 1.2. The interpretation is as follows: the odds of belonging to the joiners group computed from the proportion of households that are female headed is nearly one and a half that of households who are male headed, holding all other variables constant.

Table 5-2 shows the odds ratios and 95% confidence intervals for all significant continuous variables in Model A. Using the results of the estimated logit and odds for age in table 5-1 the estimated odds ratios for an increase in 10 years of age can be calculate as 0.54. Thus for every increase of ten years in age the likelihood of being in the joiners group decreases by about half.[115]

[115] This would be a questionable statement if it could be not expected that the additional likelihood of being a joiner for a 20 year-old as compared to a 30 year-old is similar to the additional likelihood of a 60 year-old as compared to a 50 year-old. However, it was shown that age can be treated as linear in the logit - thus this interpretation is valid over all ages.

Table 5-1: Estimated coefficients, odds ratios, standard errors, z-scores, two-tailed p-values, and 95% confidence intervals for Model A

Model A Variables Jc	Coef.	Odds Ratio	Std. Err.	Z	P>|z|	[95% C I]
Socio-economic						
age	-0.06188**	0.93999	.0005983	-103.44	0.000	-0.06305 -0.06071
gen (female)	0.13556**	1.14518	0.03254	4.17	0.000	0.07179 0.19933
depm = 1 (below)*	-0.02544**	0.97488	0.00532	-4.78	0.000	-0.03587 -0.01501
depm = 2 (above)*	-0.03039**	0.97007	0.00621	-4.89	0.000	-0.04256 -0.01822
*ref. cat. is average dependant number						
icgm = 1 (below)^	0.11218**	1.11872	0.01209	9.28	0.000	0.08848 0.13589
icgm = 2 (above)^	0.13465**	1.14414	0.02300	5.86	0.000	0.08958 0.17972
^ref. cat. is average income						
race = 1 (Asian)~	-0.82924**	0.43638	0.01863	-44.52	0.000	-0.86575 -0.79273
race = 2 (Black)~	-0.61434**	0.54100	0.01670	-36.78	0.000	-0.64707 -0.58160
race = 3 (Coloured)~	-1.42872**	0.23962	.0138872	-102.88	0.000	-1.45594 -1.40150
~ref. cat. is White						
urb (urban)	0.22973**	1.25826	0.01882	12.21	0.000	0.19284 0.26662
Claim-contribution and benefits						
tac (claims)	-0.00013**	0.99987	0.00001	-11.37	0.000	-0.00015 -0.00011
tco (contributions)	0.00054**	1.00054	0.00002	31.62	0.000	0.00051 0.00058
co2 (co-payments)	0.00019**	1.00019	0.00001	12.63	0.000	0.00016 0.00021
gp (GP benefits)	0.00045**	1.00045	0.00013	3.48	0.001	0.00020 0.00070
pub (pub.hospital benefits)	-0.00017*	0.99983	0.00007	-2.35	0.019	-0.00031 -0.00003
spe (specialist benefits)	0.00018**	1.00018	0.00001	11.84	0.000	0.00015 0.00021
prv (priv. Hospital benefits)	0.00013**	1.00014	0.00001	11.12	0.000	0.00011 0.00016
den (dental benefits)	0.00060**	1.00060	0.00006	9.34	0.000	0.00047 0.00073
opt (optical benefits)	0.00086**	1.00086	0.00007	12.15	0.000	0.00072 0.00100
par (paramedical benefits)	0.00016**	1.00016	0.00003	5.75	0.000	0.00010 0.00021
Plan classification						
risk = 1 (below)**	0.31103**	1.36483	0.02147	14.49	0.000	0.26895 0.35311
risk = 2 (above)**	-0.33860**	0.71277	0.02411	-14.05	0.000	-0.38584 -0.29135
** ref. cat. is average risk						
Interactions						
age*gen	0.01026**	1.01032	0.00082	12.59	0.000	0.00866 0.01186
age*icgm	-0.00079**	0.99921	0.00021	-3.82	0.000	-0.00120 -0.00039
age*depm	-0.00001**	0.99999	0.00000	-4.77	0.000	-0.00001 0.00000
age*opt	-0.00001**	0.99999	0.00000	-7.45	0.000	-0.00002 -0.00001
age*pub	0.00001**	1.00001	0.00000	3.22	0.001	0.00000 0.00001
depm*gen	0.04221**	1.04311	0.00510	8.28	0.000	0.03221 0.05221
depm*gp	-0.00037**	0.99963	0.00006	-6.14	0.000	-0.00048 -0.00025
icgm*race	-0.06063**	0.94117	0.00869	-6.98	0.000	-0.07766 -0.04360
icgm*den	-0.00015**	0.99985	0.00002	-6.05	0.000	-0.00020 -0.00010
race*gen	-0.00024**	0.99977	0.00003	-7.77	0.000	-0.00029 -0.00018
race*den	0.00013**	1.00013	0.00001	9.52	0.000	0.00010 0.00015
gp*tco	0.00000**	1.00000	0.00000	6.1	0.000	0.00000 0.00000
risk*gen	0.06285**	1.06487	0.01559	4.03	0.000	0.03229 0.09341
risk*icgm	0.04410**	1.04509	0.00352	12.54	0.000	0.03721 0.05100
risk*race	0.04174**	1.04263	0.00913	4.57	0.000	0.02385 0.05963
cons	1.60846**		0.03091	52.04	0.000	1.54788 1.66904
Wald chi2(37)	= 35821.34					N = 663,644
Prob > chi2	= 0.000					n = 238,561

*Note: ** significance level of 1%, * significance level of 5%*

Table 5-2: Estimated odds-ratios and 95% confidence intervals for continuous variables in Model A

Variable	Odds ratio	95% CI
age^	0.5386	0.53795 0.53922
tac (claims)^^	0.9987	0.99868 0.99872
tco (contributions)^^	1.0054	1.00538 1.00545
co2 (co-payments)^^	1.0019	1.00188 1.00192
gp (GP benefits)^^	1.0045	1.00425 1.00477
pub (pub.hospital benefits)^^	0.9983	0.99816 0.99843
spe (specialist benefits)^^	1.0018	1.00178 1.00182
prv (priv. hospital benefits)^^	1.0013	1.00128 1.00132
den (dental benefits)^^	1.0060	1.00589 1.00613
opt (optical benefits)^^	1.0086	1.00849 1.00877
par (paramedical benefits)^^	1.0016	1.00010 1.00022

Note: ^for a 10 year increase, ^^for a 10 Rand increase

The estimation of odds ratios in the presence of interaction is explained for the gender variable, controlled for age. When there is interaction between a factor and another variable, the variable odds ratio estimate also interacts with that factor. The log-odds ratio will not simplify to a single coefficient, instead it involves as many coefficients as the factor and the interaction variable have. In table 5-3 the estimated odds ratios (OR) and 95% confidence intervals (CI) for gender from Model A are calculated over different ages.

Table 5-3: Estimated odds ratios and 95% confidence intervals for gender for Model A, controlling for age

age	20	30	40	50	60	70	80
OR (gen = 1)	1.41	1.56	1.72	1.91	2.12	2.35	2.60
95% CI	1.35-1.46	1.51-1.60	1.68-1.77	1.85-1.98	2.02-2.22	2.21-2.49	2.42-2.80

The results demonstrate that the effect of gender on the odds of being a joiner increases with rising age. The increase in joining probability is significant for all ages but in particular for women 50 years and older. A female headed household that is in its reproductive years is estimated to have a joining probability one and a half times that of a household headed by a male. However, a female headed households in its retirement age (when likelihood to incur medical costs is also high) has a joining probability of more than two times that of a male headed household at the same age. The joining probability increase here can be as little as 2.4 times or as high as 2.8 times with 95% confidence. For the results presented here, however, only odds ratio calculations for significant interactions, which are of particular value to the overall results interpretation of the joiner-control model, will be discussed. For the socio-economic variables in Model A, the following results were obtained.

The probability of joining medical scheme plans decreased significantly with increasing age of principal members.[116] Females have a higher probability of joining medical scheme plans than males. The odds of joining are 1.2 times larger for female headed households than for male headed households.[117] However, the estimated odds ratios for gender, controlling for age (table 5-3), demonstrate that the effect of gender on the odds of being a joiner increases with rising age, particularly for women 50 years and older. The comparably lower odds for younger women can be partly explained by their higher likelihood to be insured through their spouses. However, the significant positive interaction gender*dependants also suggests some effect of gender on the odds of joining with an increase in the number of dependants (the odds ratios for gender, controlling for dependant number, are displayed in table A7-1 in Appendix 7). The increase in joining probability is higher for female headed households as compared to male headed households with similar numbers of dependants.[118]

Households with an above and with a below average number of dependants had a significantly lower probability of joining compared to households with an average number of dependants.[119] Results for the income group variable show that households with below and with above average income had significantly higher probabilities of joining compared to households with average incomes.[120] Race-group effects were marked, where households headed by Asian, Coloured, and Black Africans have significantly lower probabilities of joining than White Africans. Historic inequities, prejudices, and lower demand for Western health care among Non-whites were expected to explain these differences. Probability of joining is significantly higher for households in

[116] The significant interactions age*dependants, age*opt, and age*pub have very small coefficients and did not lead to considerable variations in the odds for age, controlling for dependants, for optical, or for public hospital benefits. Thus these results are not displayed here.

[117] The significant interaction gender*race, has a very small coefficient and did not lead to variations in the odds for gender, controlling for race. Similarly, the significant gender*risk interaction did not lead to variations in the odds for gender, controlling for risk plan joined. These results are not displayed here.

[118] A female household with one dependant is about one time more likely to join insurance than a similar male household. However, the joining probability of a female household rises to one and a half times of that of a male household when they both have a similar number of six dependants, all other variables being equal.

[119] The significant interaction dependants*gp has a very small coefficient, not leading to variations in the odds for dependants, controlling for GP benefits. (The results are not displayed.)

[120] The results for below average income households complies with the expectation that low income earners have a higher financial risk and thus economic incentive to seek insurance. The impact of health costs on their disposable income would be far more devastating than for wealthier households. Since above average income households also have higher odds of joining as compared to average income households, income also seems to function as an affordability constraint for a household's preference to obtain insurance coverage. The odds ratios for the models' significant interaction income*risk (in Appendix 7, table A7-2) indicate that within each risk plan type, the odds increases slightly for joining a high-risk plan at both income levels.

urban rather than rural areas. This supports the expectation that households living in urban areas have better access to health insurance (and health care).

For the contribution-claim and benefit variables the following results are displayed in table 5-2. The odds for a 10 Rand increase in claims for households that join is less than one times that of households in the control group. The probability of joining medical scheme plan coverage increases significantly for households with increasing negotiable non-essential high-end services (such as specialist and private hospital benefits). The probability of joining also increases with rising discretionary service utilisation, i.e rising dental, optical, and paramedical benefits. This is also true for the essential service type of GP benefits. However, joining probability increases with decreasing essential public hospital benefits. A household's probability of joining medical scheme plans increases with increasing contributions and co-payments.

Finally, the results for the plan classification variable show that the probability of joining increases with households' probabilities of entering low-risk plans as compared to households entering average-risk plans. The odds of joining are 1.4 times larger for households who joined below average risk plans than for households who joined average-risk plans. However, joining probability decreases with the decreasing probability of households entering plans with above average risk profiles. The odds of joining, for households that joined high-risk plans, are only two-thirds that of similar households joining average-risk plans.[121] The next section will present the estimation results of the leavers-control model, Model B.

5.2.2 Model B results: leavers-control group

Model B tested for differences between the risk profiles of leavers (households that leave medical scheme plans during any time during the observation period) and the control group. The dependant variable for the Model B is lc, where lc=1 for households that left at any point during the observation period and zero otherwise. On the next page table 5-4 displays Model B's coefficients, odds ratios, semi-robust standard errors, z-scores, two-tailed *p*-values, and 95% confidence intervals for the population average model, with exchangeable correlation. The 684,939 observations that entered Model B were clustered over 262,349 households. The minimum of observations per household is 1 (year) and the

[121] The risk*income interaction was significant. The odds ratios for risk, controlling for income, are in table A7-3 (Appendix 7). While within each income group the odds of joining increased for high/low risk plans, compared to average plans, the odds increase was highest for above average incomes joining a low risk plan.

Table 5-4: Estimated coefficients, odds ratios, standard errors, z-scores, two-tailed p-values, and 95% confidence intervals for Model B

Model B Variables lc	Coef.	Odds Ratio	Std. Err.	z	P>|z|	[95% C I]
Socio-economic						
age	-0.03612**	0.96453	0.00064	-56.08	0.000	-0.03738 -0.03485
gen (female)	0.68371**	1.98122	0.02963	23.07	0.000	0.62563 0.74179
depm = 1 (below)*	-0.06332**	0.93864	0.00948	-6.68	0.000	-0.08191 -0.04473
depm = 2 (above)*	-0.05722**	0.94439	0.01702	-3.36	0.001	-0.09057 -0.02386
*ref. cat. is average dependant number						
icgm = 1 (below)^	0.16825**	1.18323	0.01195	14.07	0.000	0.14482 0.19168
icgm = 2 (above)^	0.22438**	1.25155	0.02366	9.48	0.000	0.17800 0.27076
^ref. cat. is average income						
race = 1 (Asian)~	-0.99478**	0.36980	0.02053	-48.45	0.000	-1.03503 -0.95454
race = 2 (Black)~	-1.10995**	0.32958	0.02573	-43.14	0.000	-1.16038 -1.05952
race = 3 (Coloured)~	-1.27728**	0.27879	0.03342	-38.22	0.000	-1.34278 -1.21178
~ref. cat. is White						
urb (urban)	0.21248**	1.23674	0.01809	11.75	0.000	0.17702 0.24793
Claim-contribution and benefits						
tac (claims)	0.00005**	1.00005	0.00001	3.9	0.000	0.00003 0.00008
tco (contributions)	0.00024**	1.00024	0.00002	10	0.000	0.00019 0.00029
co2 (co-payments)	0.00027**	1.00027	0.00005	5.59	0.000	0.00018 0.00036
gp (GP benefits)	0.00093**	1.00093	0.00004	22.22	0.000	0.00085 0.00101
pub (pub.hospital benefits)	-0.00015*	0.99985	0.00007	-2.01	0.045	-0.00029 0.00000
spe (specialist benefits)	-0.00006**	0.99994	0.00002	-3.73	0.000	-0.00010 -0.00003
prv (priv. hospital benefits)	-0.00008**	0.99992	0.00001	-5.95	0.000	-0.00010 -0.00005
den (dental benefits)	0.00111**	1.00111	0.00006	19.75	0.000	0.00100 0.00122
opt (optical benefits)	0.00198**	1.00198	0.00008	23.86	0.000	0.00182 0.00214
par (paramedical benefits)	0.00008*	1.00008	0.00004	2.23	0.026	0.00001 0.00015
Plan classification						
risk = 1 (below)**	-0.08470**	0.91879	0.01501	-5.64	0.000	-0.11412 -0.05528
risk = 2 (above)**	0.21934**	1.24526	0.01864	11.77	0.000	0.18281 0.25587
** ref. cat. is average risk						
Interactions						
age*gen	-0.01187**	0.98820	0.00071	-16.83	0.000	-0.01325 -0.01049
age*depm	0.00067**	1.00067	0.00021	3.21	0.001	0.00026 0.00107
age*icgm	-0.00250**	0.99751	0.00019	-13.22	0.000	-0.00287 -0.00213
age*race	0.00759**	1.00762	0.00026	29.63	0.000	0.00709 0.00810
age*den	-0.00002**	0.99998	0.00000	-14.67	0.000	-0.00002 -0.00002
age*opt	-0.00002**	0.99998	0.00000	-11.91	0.000	-0.00002 -0.00002
age*co2	-0.00001**	0.99999	0.00000	-6.19	0.000	-0.00001 0.00000
age*pub	0.00000*	1.00000	0.00000	2.5	0.012	0.00000 0.00001
age*tac	0.00000**	1.00000	0.00000	2.88	0.004	0.00000 0.00000
gen*icgm	-0.01372**	0.98638	0.00471	-2.91	0.004	-0.02294 -0.00449
race*tco	-0.00019**	0.99981	0.00001	-17.63	0.000	-0.00021 -0.00017
icgm*race	0.02922**	1.02965	0.00206	14.17	0.000	0.02518 0.03326
icgm*urb	-0.05555**	0.94596	0.00889	-6.25	0.000	-0.07298 -0.03812
opt*tac	0.00000**	1.00000	0.00000	-2.85	0.004	0.00000 0.00000
risk*icgm	0.03389**	1.03447	0.00372	9.11	0.000	0.02659 0.04118
cons	0.87973**		0.03334	26.38	0.000	0.81438 0.94508
Wald chi2(37)	= 20066.5					N = 684,939
Prob > chi2	= 0.000					n = 262,349

*Note: ** significance level of 1%, * significance level of 5%*

maximum is 4 years, with an average of 2.6 observations per household. Model B's covariates are all at least significant at the five percent level. The following estimated within household working correlation matrix R can be reported:

	c1	c2	c3	c4
r1	1.0000			
r2	0.8455	1.0000		
r3	0.8455	0.8455	1.0000	
r4	0.8455	0.8455	0.8455	1.0000

Table 5-5 displays the calculated odds ratios and 95% confidence intervals for all significant continuous variables in Model B. The estimated odds ratio for an increase in 10 years of age is 0.7. This result can be interpreted as meaning that for every increase of ten years in age the likelihood of being in the leavers group decreases by about one third. Since age can be treated as linear in the logit, the interpretation is valid over all ages.

Table 5-5: Estimated odds-ratios and 95% confidence intervals for continuous variables in model B

Variable	Odds ratio	95% CI
age^	0.6968	0.69596 0.69771
tac (claims)^^	1.0005	1.00048 1.00052
tco (contributions)^^	1.0024	1.00236 1.00244
co2 (co-payments)^^	1.0027	1.00261 1.00280
gp (GP benefits)^^	1.0093	1.00926 1.00942
pub (pub.hospital benefits)^^	0.9985	0.99836 0.99863
spe (specialist benefits)^^	0.9994	0.99936 0.99943
prv (priv. hospital benefits)^^	0.9992	0.99918 0.99921
den (dental benefits)^^	1.0111	1.01104 1.01128
opt (optical benefits)^^	1.0198	1.01984 1.02016
par (paramedical benefits)^^	1.0008	1.00072 1.00088

Note: ^for a 10 year increase and ^^for a 10 Rand increase

First, the results for the socio-economic variables are presented. Probability of leaving medical scheme plan coverage decreases significantly with increasing age of principal members.[122] Females have a higher probability of leaving insurance coverage. The odds of leaving are nearly two times larger for female headed households than that for male headed households. However, the calculated odds ratios for gender, controlling for age, displayed in the Appendix 7 in table A7-4, demonstrate that the effect of gender on the odds of being a leaver decrease with rising age, particularly for women younger than 40.[123]

[122] The calculated odds ratios for the significant interactions age*den, age*opt, age*pub, age*co2, and age*tac, controlling age for dental, optical and public hospital benefits, co-payments, and claims are not displayed here. Because these interactions were extremely small they add not further interpretable information.

[123] Female headed households in the reproductive years have a leaving probability of 1.5, but in the retirement age of only 0.8 times that of households headed by similarly aged males.

Households with an above and with a below average number of dependants have a significantly lower probability of leaving insurance as compared to households with average numbers of dependants.[124] The results for income show that households with below and with above average incomes have a significantly higher probability of leaving insurance as compared to households with average incomes.[125] Race-group effects were significant, with households headed by Asian, Coloured, and Black Africans having a significantly lower probability of leaving insurance coverage than White Africans. The explanation for these results might lay in unequal access for Non-whites to health insurance options outside the medical scheme market. Probability of leaving is significantly higher for households in urban, as opposed to rural, areas, indicating better access for urban households to different insurance options.[126]

For the claim and benefits variables the odds ratios are displayed in table 5-5. The odds of a 10 Rand increase in claims for households that leave is more than one times that of a household in the control group. The probability that households leave medical scheme plans increases significantly with decreasing negotiable non-emergency services: specialist and private hospital benefits. This is also true for the essential service type of public hospital benefits, while the probability of leaving increases with increasing essential service utilisation of GP benefits.[127] The probability of leaving increases with rising discretionary service utilisation, i.e rising dental, optical, and paramedical benefits. Probability of leaving increases with increasing contributions and co-payments.

For the plan classification variable the following results were obtained. The probability of leaving decreases with households' probability of leaving low-risk plans

[124] The calculated odds ratios for dependants, controlling for age are displayed in Appendix 7, table A7-5. The increase in leaving probability is significant for older households near retirement age with above average dependant numbers. Households with above average dependant numbers of age older than 60 have a one time higher leaving likelihood than that of households at the same age with average number of dependants.
[125] The fact that below average income households have higher odds of leaving than households with average incomes, holding all other variables equal, matches the expectation that low income presents a strong affordability constraint. However, it contradicts the expectation that low-income earners have a higher economic incentive to keep insurance. Thus it seems very likely that, if low-income earners leave, they seek insurance somewhere else. The significant interaction age*income (Appendix 7, table A7-6) showed decreasing odds for leaving with age. The odds ratio for the income*risk interaction (Appendix 7, table A7-7) showed that households, regardless their income, have increasing odds of leaving high-risk plans. Thus income functions as an affordability constraint for insurance coverage in general and for alternative choices.
[126] Decreasing odds that urban households leave (calculated for the significant interaction urban*income in Appendix 7, table A7-8) suggests that urban households with high income have a lower leaving probability.
[127] However, the public hospital benefit coefficient only became significant at the 95% level after controlling for age. Also, in the section describing the data subgroups it was shown that leavers had on average the highest public hospital benefits among all mover groups, exceeding the mean of the control group significantly. Leaving probability increases with increasing GP benefits. However, it was assumed that in most health care systems GPs function as gatekeepers for high-end services.

compared to households leaving average-risk plans. However, leaving increases with the probability of households leaving plans with above average risk profiles.[128]

The next section will present the estimation results of the switchers-control model, Model C.

5.2.3 Model C results: switchers-control group

Model C tested for differences between the risk profiles of switchers (households that switch medical scheme plans during any time during the observation period) and the control group. The dependant variable for Model C is sc, with sc=1 for households that switched a plan at any point during the observation period and zero otherwise. Switchers were coded in the data either if households joined or left a plan within a medical scheme. It could be not distinguished where a switcher that joined came from or where switchers that left went. However, of the about 31,000 switcher observations, more than 90% are from switchers that joined. Because of the small number of switchers that leave, both types of switchers had to be pooled for the statistical model analysis. For the interpretation it is of importance to keep in mind that these results mainly concern switchers that joined a new plan.

On the next page table 5-6 displays Model C's coefficients, odds ratios, semi-robust standard errors, z-scores, two-tailed *p*-values, and 95% confidence intervals for the population average model, with exchangeable correlation. The 532,010 observations that entered Model C were clustered over 177,981 households. The minimum of observations per household is 1 (year) and the maximum 4 years, with an average of 3 observations per household. Model C's covariates are all significant at the five percent level, except the below average dependants number, the above average income group, and paramedical benefits. The following estimated within household working correlation matrix R can be reported:

	c1	c2	c3	c4
r1	1.0000			
r2	0.7715	1.0000		
r3	0.7715	0.7715	1.0000	
r4	0.7715	0.7715	0.7715	1.0000

[128] The significant interaction risk*income leads to calculated odds (Appendix 7, table A7-9) that increase within each income group. However, the odds increase is higher for above average income households to leave a high-risk plan rather than an average-risk plan.

Table 5-6: Estimated coefficients, odds ratios, standard errors, z-scores, two-tailed p-values, and 95% confidence intervals for Model C

Model C Variables sc	Coef.	Odds Ratio	Std. Err.	z	P>|z|	[95% C I]
Socio-economic						
age	-0.00671**	0.99331	0.00101	-6.66	0.000	-0.00869 -0.00474
gen (female)	0.24360**	1.27583	0.06394	3.81	0.000	0.11828 0.36891
depm = 1 (below)*	0.01540	1.01552	0.01187	1.3	0.195	-0.00787 0.03866
depm = 2 (above)*	-0.05222**	0.94912	0.01166	-4.48	0.000	-0.07508 -0.02936
*ref. cat. is average dependant number						
icgm = 1 (below)^	0.08251**	1.08601	0.01188	6.94	0.000	0.05922 0.10580
icgm = 2 (above)^	0.00203	1.00203	0.01718	0.12	0.906	-0.03165 0.03570
^ref. cat. is average income						
race = 1 (Asian)~	0.13200**	1.14111	0.03299	4	0.000	0.06735 0.19665
race = 2 (Black)~	-0.65239**	0.52080	0.04349	-15	0.000	-0.73763 -0.56715
race = 3 (Coloured)~	-1.27134**	0.28045	0.04804	-26.46	0.000	-1.36550 -1.17719
~ref. cat. is White						
urb (urban)	0.11245*	1.11901	0.04903	2.29	0.022	0.01636 0.20854
Claim-contribution and benefits						
tac (claims)	-0.00007**	0.99993	0.00002	-2.93	0.003	-0.00012 -0.00002
tco (contributions)	-0.00063**	0.99937	0.00013	-4.96	0.000	-0.00088 -0.00038
co2 (co-payments)	0.00031**	1.00031	0.00003	8.83	0.000	0.00024 0.00037
gp (GP benefits)	0.00103**	1.00104	0.00027	3.8	0.000	0.00050 0.00157
pub (pub.hospital benefits)	-0.00085**	0.99915	0.00017	-4.85	0.000	-0.00119 -0.00050
spe (specialist benefits)	0.00008*	1.00008	0.00003	2.28	0.022	0.00001 0.00014
prv (priv. hospital benefits)	0.00008**	1.00008	0.00003	3.21	0.001	0.00003 0.00013
den (dental benefits)	0.00022**	1.00022	0.00005	4.36	0.000	0.00012 0.00032
opt (optical benefits)	-0.00077**	0.99923	0.00011	-6.98	0.000	-0.00098 -0.00055
par (paramedical benefits)	-0.00008	0.99992	0.00006	-1.28	0.200	-0.00021 0.00004
Plan classification						
risk = 1 (below)**	0.42828**	1.53462	0.08650	4.95	0.000	0.25875 0.59782
risk = 2 (above)**	-0.56429	0.56877	0.17065	-3.31	0.001	-0.89875 -0.22983
** ref. cat. is average risk						
Interactions						
age*gen	-0.00987**	0.99018	0.00139	-7.12	0.000	-0.01258 -0.00715
age*tco	0.00002**	1.00002	0.00000	6.14	0.000	0.00001 0.00002
gen*icgm	0.04919**	1.05042	0.01299	3.79	0.000	0.02372 0.07466
gen*gp	0.00093**	1.00093	0.00022	4.22	0.000	0.00050 0.00137
icgm*race	-0.08189**	0.92137	0.00586	-13.98	0.000	-0.09337 -0.07041
icgm*gp	-0.00085**	0.99915	0.00017	-5.05	0.000	-0.00118 -0.00052
race*spe	-0.00004	0.99996	0.00002	-1.81	0.071	-0.00009 0.00000
race*tco	-0.00015**	0.99985	0.00003	-4.38	0.000	-0.00022 -0.00008
urb*pub	0.00069**	1.00069	0.00018	3.78	0.000	0.00033 0.00104
risk*age	-0.00601**	0.99401	0.00086	-6.97	0.000	-0.00770 -0.00432
risk*gen	0.25842**	1.29489	0.02626	9.84	0.000	0.20697 0.30988
risk*icgm	0.02203*	1.02227	0.00894	2.46	0.014	0.00451 0.03955
risk*urb	0.31010**	1.36356	0.07107	4.36	0.000	0.17080 0.44940
risk*opt	0.00079**	1.00079	0.00010	7.73	0.000	0.00059 0.00099
risk*tco	0.00031**	1.00032	0.00005	6.81	0.000	0.00022 0.00041
cons	-1.88824**		0.07114	-26.54	0.000	-2.02767 -1.74882
Wald chi2(37)	= 5431.19					N = 532,010
Prob > chi2	= 0.000					n = 177,981

*Note: ** significant level of 1%, * significance level of 5%*

In table 5-7 the calculated odds ratios and 95% confidence intervals for all significant continuous variables in Model C are displayed. The calculated odds ratio for a 10 year increase in age is 0.93. The interpretation is that for every increase of ten years in age the likelihood of being in the joiners group decreases by about 7%.[129]

Table 5-7: Estimated odds-ratios and 95% confidence intervals for continuous variables in Model C

Variable	Odds ratio	95% CI	
age^	0.9331	0.91677	0.95379
tac (claims)^^	0.9993	0.99926	0.99933
tco (contributions)^^	0.9937	0.99346	0.99397
co2 (co-payments)^^	1.0031	1.00305	1.00316
gp (GP benefits)^^	1.0135	1.00981	1.01089
pub (pub.hospital benefits)^^	0.9915	0.99120	0.99186
spe (specialist benefits)^^	1.0008	1.00074	1.00085
prv (priv. hospital benefits)^^	1.0008	1.00074	1.00085
den (dental benefits)^^	1.0022	1.00210	1.00230
opt (optical benefits)^^	0.9923	0.99211	0.99254
par (paramedical benefits)^^	Not significant		

Note: ^for a 10 year increase, ^^for a 10 Rand increase

First, the following results for the socio-economic variables were obtained. Probability of switching medical scheme plan coverage decreases significantly with increasing age of the principal member.[130] Females have a higher probability of switching insurance coverage. The odds of switching are 1.3 times larger for female headed households than for male headed households. However, the calculated odds ratios for gender, controlling for age, (Appendix 7, table A7-10) demonstrate that the effect of gender on the odds of being a switcher decreases with rising age, particularly for women between 20 and 40 years of age.[131] Also, the significant interaction gender*risk leads to a substantial odds ratio increase for female headed households switching to high-risk plans (rather than for switching to low-risk plans as compared to switching to an average plan).[132]

Households with an above average number of dependants have a significantly lower probability of switching insurance as compared to households with an average number of dependants. Households with below average incomes have a significantly

[129] Since age can be treated as linear in the logit the interpretation is valid over all ages.
[130] The age*claim interaction had extremely small coefficients near zero. The calculated odds ratios for the significant interaction of age*tco, with age controlled for contributions did not contain further information for this interpretation and thus the results are not displayed here.
[131] While female headed households in their reproductive years have an estimated switching probability of more than one time that of households headed by similarly aged males, female headed households in their retirement age have a switching probability of only two thirds that of male headed households.

higher probability of switching insurance plans as compared to households with average incomes, while above average income households have a positive but not significant coefficient for switching plans compared to average income households.[133] Significant race-group effects were detected, with households headed by Coloured and Black Africans having a significantly lower probability of switching insurance coverage than White Africans.[134] Probability of switching is significantly higher for urban households than for rural households, supporting the hypothesis of better access to insurance options and health services for urban households.[135]

In Model C the claim and benefits variables displayed the following results. The odds for a 10 Rand increase in claims for a household that switches is less than that of a household in the control group (table 5-7). The probability that households switch medical scheme plans increases significantly with increasing negotiable non-emergency service utilisation of specialist and private hospital benefits. Switching also increases with rising discretionary service utilisation of dental benefits and decreasing optical benefits (paramedical benefits did not predict probability of switching). The switching probability decreases with essential public hospital benefits but increases with GP benefits. Probability of switching insurance increases with decreasing contributions and increasing co-payments.

For the plan classification variable the probability of switching increases with the probability that households switch to low-risk plans as compared to households switching to average-risk plans. The odds of switching are 1.5 times larger for households that switch to below average risk plans than those of households switching to average-risk plans. Switching probability decreases with the probability of a household switching to a plan with an above average risk profile. The odds of switching, for households that switch

[132]Table A7-11 in Appendix 7 shows that female headed households switching to high-risk plans are over two times more likely to switch than male households. Female headed households switching to low-risk plans are only one and a half time more likely to switch compared to male headed households.

[133] The first coefficient matches the expectation that low income earners have a higher economic incentive to keep insurance, even if it means switching to other plans. Again, the impact of health costs on their disposable income would be far more devastating than for wealthier households. The sign of the second, above average income coefficient supports the hypothesis that income functions as an affordability constraint.

[134] However, the odds of switching for households headed by Asian Africans are higher than for households headed by White Africans. The interaction race*income is significant in Model C and the calculated odds ratios for race, controlling for income, displayed in Appendix 7, table A7-12, reveal decreases in the switching probability for Asian, Black, or Coloured households within the below average income group as compared to White households. Particularly noticeable is the decrease for Asian households' switching probability after controlling for income. It becomes nearly equal to that of similar White households. Prevailing economic inequities might be here a valid interpretation.

to above average risk plans, are only half of those of similar households switching to average-risk plan.[136]

Before the result of the three models could be interpreted in regard to evidence for adverse selection and/or risk selection they were assessed for their model fit. This is the last step of each statistical analysis based on model building. The next section will shortly summarise the results of the model fit assessment for the Models A, B, and C.

5.3 Results of model fit assessment

The purpose of this section is to present the results of the model fit assessment for the three models used in this thesis' empirical analysis. Several standard methods and diagnostic statistics for the overall fit of logistic regression models are available, which range from summary measures of goodness-of-fit to logistic regression diagnostics. However, diagnostic statistics for the overall fit of a logistic regression model have not yet been extended for the use in model assessments for correlated data models (Hosmer and Lemeshow 2000).

Hosmer and Lemeshow (2000) suggest that in such a case the analysis should be approximated with standard methods developed for normal uncorrelated data, rather than the inferior option of performing no model checking at all. This means, in the case of this thesis' analysis, that instead of fitting the marginal models (for panel data) and assessing their model fit, similar logit models (designed for cross-sectional data) should be fitted and assessed for their fit. Or in other words, simple logit models should be fitted and then be assessed for their model fit using correlated data, but assuming that the observations are in fact not correlated. However, a clear disadvantage of this procedure will be that the then fitted and assessed models will not consider the typical panel data correlation. Thus the model fit assessment of these proxy models has to be interpreted with caution. Applying these methods of goodness-of-fit and logistic regression diagnostics will merely indicate the fit of the actual marginal models, but may not stand up to the standards of precise measurement.

[135] The significant interaction urban*risk demonstrates that the odds of switching increase for urban households, no matter to which plan they switch (Appendix 7, table 7-13). The odds increase is significantly higher for switching to high-risk plans, compared to similar urban households switching to low-risk plans.
[136] The estimated coefficient for the risk*age interaction was significant and the effects of risk on the odds of switching, controlled for age (presented in Appendix 7, table A7-14) show that the decrease in the odds of switching with rising age is significant, in particular, for switchers to low-risk plans.

In order to obtain the results discussed in this model assessment section all three Models, A, B, and C, were estimated using the LOGIT command in STATA™ for normal cross-sectional data (i.e. assuming uncorrelated data). The ROBUST option allowed the specification of clusters (here over household IDs) such that within the clusters the assumption of independence was relaxed.

There are several standard approaches for assessing the fit of models, which range from the evaluation of overall measures of fit to the examination of individual components of the summary statistics, and logistic regression diagnostic, often with graphs. Summary measures of goodness-of-fit are provided in most software packages with the output for any fitted model. Although summary statistics provide only limited information on the fit of model components, they can indicate interesting deviations from the fit. The complete model fit assessment is presented in detail in Appendix 8. Here only a short summary will be presented.

The summary measures of goodness-of-fit, like the Pseudo-R-Square, the Pearson goodness-of-fit chi-square statistics, and the Hosmer-Lemeshow goodness-of-fit test did not show a convincing fit for the proxy models. This could be expected, considering the inappropriateness of these proxy models for the data at hand. However, the results of the classification tables and ROC (Receiver Operating Characteristics) curves for the proxy models indicated acceptable to good classification and discrimination. Generally, overall goodness-of-fit measures and classification tables are only one part of model fit assessments. They provide single numbers that summarise a considerable amount of information. Given the model, the context and the objective, some measures performed better, some worse (see also the discussion in chapter 6).

For a better judgement of a model's fit, logistic regression diagnostics are performed in order to see if a fit was supported over the entire covariate pattern. The examination of covariate patterns is very important, since the previously discussed summary statistics, based on the Pearson chi-square residuals, only summarise the agreement between observed and fitted values in a single and hard to assess number. These diagnostic regression techniques are nearly exclusively based on the visual assessment of graphs (Hosmer and Lemeshow 2000).

Components of the residual sum-of squares are key quantities for logistic regression diagnostics. These are first, the residuals for each covariate pattern, the hat

matrix, and the leverage values derived from it.[137] The plots of leverage values versus Pearson residuals for all three proxy models (in Appendix 8) demonstrate that only a very small number of covariate patterns deviated, which indicates good model fit.

Another diagnostic statistic examined the decrease in the Pearson chi-square goodness-of-fit statistic ($\Delta\chi^2$) and the effect on the value of the estimated coefficients ($\Delta\hat{\beta}$) caused by deleting an observation (and all others sharing the covariate pattern).[138] These diagnostic techniques can easily identify covariate patterns that are poorly fit (large values of $\Delta\chi^2$) and those that have a great deal of influence on the values of estimated parameters (large values of $\Delta\hat{\beta}$).

Among the selection of various types of plots, each focusing on a different aspect of fit, three were chosen here. The first two are the plots of $\Delta\chi^2$ and of $\Delta\hat{\beta}$ versus the estimated logistic probability. The third is a plot of $\Delta\chi^2$ versus the estimated logistic probability where the size of the plotting symbol is proportional to the size of $\Delta\hat{\beta}$. These three plots for the proxy models, presented in the Appendix 8, show only a very small number of deviating covariate patterns and indicate very good model fit.

The limitations of these summary measures of goodness-of-fit and logistic regression diagnostics are discussed in chapter 6. However, in this case all goodness-of-fit measures and logistic regression diagnostics were applied to the simple logistic models, which only functioned as proxies for the actual models in order to make up for the lack of fit assessment tools for models with correlated data. After this thorough assessment it can be concluded that even the proxy models are well fit. That leads to the conclusion that the actual models used, which are more appropriate for the data, fit very well.

The model fit assessment results were the final results for this thesis' empirical analysis. The next section will interpret the model estimation results.

[137] The hat matrix is the matrix that provides the fitted values as the projection of the outcome variable into the covariate space. The diagonal elements of the hat matrix are called the leverage value. They are proportional to the distance from a covariate to the mean of the data (Hosmer and Lemeshow 2000).
[138] For the mathematical foundation of these diagnostics see Hosmer and Lemeshow (2000) p.173.

5.4 Interpretation of model estimation results

The purpose of this section is to interpret the estimation results of the presented logistic regression models. This interpretation will focus on evidence for adverse selection and risk selection, which, as hypothesised, would lead to a lack of risk pooling in South Africa's unregulated medical scheme market. As discussed on the previous chapters, this does not exclude the possibility of a separating equilibrium, which might indeed be efficient but yet not equitable. This thesis's data and the empirical analysis design do not lend themselves to an explicit test for the existence of a separating equilibrium, however, this possible interpretation of the results will be discussed. This section will first, evaluate the empirical analysis results within the group method framework. Then the interpretations for Model A of joiners and the control group will be discussed, followed by Model B, of leavers and the control group, and finally Model C, of switchers and the control group.

5.4.1 The results within the group method framework

The group method aimed to derive evidence for adverse selection and risk selection, by defining population subgroups within the insured population - households that enter, exit, and switch insurance plans compared to a control group of the constantly insured. The selection processes were categorised into several subtypes in order to specify the expected observations for households' risk profiles under each type of adverse selection and risk selection. This yielded two matrix-style tables, one for each selection type, specifying, among other things, the expected empirical observations for each group. The empirical analysis results will be synthesised on the next pages with these two tables, highlighting coinciding expected and actual observations.

Evidence for adverse selection, summarised with table 5-8 on the next page, is rather limited. Among adverse selection types one and two, only the more likely entry of females and thus more frequent female membership among joiners and switchers, consistently indicates adverse selection. Generally, a more frequent choice of insurance entrance among female headed households was expected to be one (among several other) indicating high-risk factor.[139] However, aside from this variable only higher non-essential benefits supports the adverse selection hypothesis. Other, more substantial, components

[139] However, it was pointed out in the gender variable expectation and results interpretation that gender biases might occur in employment and compensation. For this reason it was not deemed a very good risk indicator for the conclusion of the existence of evidence regarding selection processes.

of the household profiles did not indicate the preferred entry of high risks into the medical scheme market or specifically high-risk plans. Thus from such a limited number of not very strong risk indicators adverse selection of type one or two cannot be concluded.

However, table 5-8 shows more evidence for adverse selection of type three, where low risks discontinue insurance coverage in the observed medical scheme plans (presumably giving up insurance coverage altogether or joining a for-profit insurer).

Table 5-8: Concluded adverse selection types

Adverse Selection	Type 1 Choice of insurance coverage by high risks	Type 2 Choice of certain plans of insurance coverage by high risks	Type 3 Choice of discontinued insurance coverage by low risks
MANIFESTATION	Entrance of High Risks	**A.** Entrance of High Risks into High-Risk Plans **B.** Switch of High Risks to High-Risk Plans	**A.** Exit of Low Risks **B.** Exit of Low Risks from High-Risk Plans
MECHANISMS	ADVERSE SELECTION OF HIGH RISKS	ADVERSE SELECTION OF HIGH RISKS	ADVERSE SELECTION OF LOW RISKS
EFFECTS	High Risks join on average more	**A.** High Risks join on average more high-risk plans **B.** High risks switch on average more to high-risk plans	**A.** Low risks leave on average more **B.** Low risks leave on average more high-risk plans
AFFECTED GROUPS	JOINERS	**A.** JOINERS **B.** SWITCHERS	**A.** LEAVERS **B.** LEAVERS
EXPECTED EMPIRICAL OBSERVATIONS AND COINCIDING **ACTUAL OBSERVATIONS**	On average Joiners: 1. Are older, **female**, non-single 2. Have higher claims, lower contribution-claim ratios 3. Have higher essential and **higher non-essential benefits**	A On average Joiners: 1. Are older, **female**, non-single 2. Have higher claims, lower contribution-claim ratios 3. Have higher essential and **higher non-essential benefits** Join high-risk plans B On average Switchers: 1. Are older, **female**, non-single 2. Have higher claims, lower contribution-claim ratios 3. Have higher essential and **higher non-essential benefits** Switch to high-risk plans	A On average Leavers: 1. Are **younger**, male, single 2. Have lower claims, higher contribution-claim ratios 3. Have **lower essential, lower non-essential service utilisation and higher discretionary benefits** B On average Leavers: 1. Are **younger**, male, single 2. Have lower claims, higher contribution-claim ratios 3. Have **lower essential lower non-essential service utilisation and higher discretionary benefits** **Leave from high-risk plans**

Table 5-9: Concluded risk selection types

Risk Selection	**Type 1** **Discourage high-risk insurance applicants**	**Type 2** **Attract low risks to certain plans of insurance coverage**	**Type 3** **Discourage high risks to continue insurance coverage or switch them to high-risk plans**
MANIFESTATION	Restricted Entrance of High Risks	**A.** Entrance of Low Risks into Low-Risk Plans **B.** Switch of Low Risks to Low-Risk Plans	**A.** Exit of High Risks (from High-Risk Plans **B.** Switch of High Risks to High-Risk Plans
MECHANISMS	UNFAVOURABLE RISK SELECTION	FAVOURABLE RISK SELECTION/CREAM-SKIMMING	A. DUMPING B. UNFAVOURABLE RISK SELECTION
EFFECTS	High risks join on average less	**A.** Low risks join on average more low risk plans **B.** Low Risks switch on average more to low risk plans	**A.** High Risks leave on average more (high risk plans) **B.** High risks switch on average more to high risk plans
AFFECTED GROUPS	JOINERS	**A.** JOINERS **B.** SWITCHERS	**A.** LEAVERS **B.** SWITCHERS
EXPECTED EMPIRICAL OBSERVATIONS AND COINCIDING **ACTUAL OBSERVATIONS**	On average Joiners: 1. Are **younger**, male, single 2. Have **lower claims, higher contribution-claim ratios** 3. Have **lower essential**, lower non-essential **and higher discretionary benefits**	A. On average Joiners: 1. Are **younger**, male, single 2. Have **lower claims, higher contribution-claim ratios** 3. Have **lower essential**, lower non-essential **and higher discretionary benefits** **Join low-risk plans** B. On average Switchers: 1. Are **younger**, male, single 2. Have **lower claims**, higher contribution-claim ratios 3. Have **lower essential**, lower non-essential and higher discretionary benefits **Switch to low-risk plans**	A. On average Leavers: 1. Are older, **female**, no-single 2. Have **higher claims, higher contributions, lower contribution-claim ratios** 3. Have higher essential and higher non-essential benefits **Leave high-risk plans** B. On average Switchers: 1. Are older, **female**, no-single 2. Have higher claims, higher contributions, **lower contribution-claim ratios** 3. Have higher essential and **higher non-essential benefits** Switch to high-risk plans

Several components of the households' risk profiles, like age, benefit characteristics and the type of plan that is more frequently exited indicate that low risks leave. Specifically it shows that these low risks, are more likely to leave high-risk plans, i.e. risk pools with a higher a presence of high risks.

Thus it seems that adverse selection in the form of low risks discontinuing insurance coverage in the South African medical schemes can be concluded. However, while discussing the expectations, it was pointed out that this type of adverse selection

might be induced by external favourable risk selection, depending on the degree of competition between insurers. Two facts speak for the interpretation of these last results as evidence for external favourable risk selection. First, no considerable evidence of adverse selection *into* the medical scheme plans could be found. This means the typical adverse selection cycle, where high risks disproportionately enter specific high-risk plans and in turn low risks give up insurance coverage, cannot be confirmed. Second, the South African medical scheme market in the 1990s was surely a contested health insurance market and the better explanation is that low risks left due to aggressive underwriting by the medical scheme plan competitors.

Altogether, much stronger evidence for risk selection of all three types could be found, as summarised in table 5-9 on the previous page. For example, for risk selection type one and type 2A, joiners clearly display a low risk profile. They are younger and have lower claims, a contribution-claim ratio over one, lower essential benefits, and higher discretionary benefits. Also, joiners predominantly enter low-risk plans. All these factors are strong indicators for low risk. This allows the interpretation of these results as overwhelming evidence for risk selection in the form of discouraging the entry of high risks and cream-skimming of low risks. Switchers also display a low risk profile and predominantly switch to low-risk plans, confirming the cream-skimming hypothesis. Finally, risk selection also works on the side of market exit, where high risks were dumped. Leavers' risk profiles display, in particular, high claims and contribution-claim ratios under one and a tendency to leave high-risk plans. These results can be interpreted as evidence for risk selection type three, where high risks are pressured to leave insurance coverage.

This summary of the results interpretation predominantly demonstrates evidence for risk selection in the South African medical scheme data during the time of observation. Unfavourable and favourable risk selection by medical scheme plans led to the observation of cream-skimming and dumping. The medical scheme market was also contested at the time of observation - external for-profit insurers competed aggressively for low risks from the medical scheme plan segment. The next three sections will discuss the results interpretations for Model A, Model B and Model C.

5.4.2 Model A results interpretation: joiners-control group

First, among the socio-economic variables in Model A, age delivered very good evidence for selection. Younger people are more likely to join medical scheme plans, because their lower risk for ill-health makes them a target for favourable risk selection (cream-skimming) by insurers. Younger people might also have a higher joining probability because they tend to enter professional life and take out health insurance. However, the decrease in joining probability with age was interpreted as unfavourable risk selection by insurers against the elderly in the insurance market where older persons have poorer access to health insurance coverage. Given that older households have a higher likelihood of ill-health and thus health care costs, it is unlikely that they should express a lower joining preference.

Interestingly, the results for the variables gender and number of dependants also could be interpreted as evidence for selection processes. For example, females have a higher probability of joining medical scheme plans than males and older females are more likely to join, as are females with several dependants. Given that females have a higher expected risk of using health care resources, particularly, when older, and that females are more likely to seek insurance with further dependants compared to similar males, it can be assumed that adverse selection in terms of females joining health insurance takes place.

Households with an above average number of dependants have a higher likelihood of health related costs, and a lower joining probability. This result is interpreted as evidence for unfavourable risk selection practised by the insurers against large households, because the later seem to have comparably more difficulties in obtaining medical scheme coverage. However, households with below average numbers of dependants also join less frequently. Presumably, households with no dependants have less economical incentive to pay for medical scheme plan coverage that is rarely used.
As expected, the results for the variables income, race, and residency cannot be interpreted as evidence for adverse selection and risk selection.

Contribution-claim and benefit variables that are some function of prior utilisation were expected to deliver the best evidence for selection. The results and interpretations in regard to evidence for selection processes are very good for the variables of claims and benefits. The negative coefficient for claims can be interpreted as evidence for favourable risk selection of low risks by the insurer among insurance aspirants that results in lower claims among new recruits. Given that households with high claims have a higher likelihood of ill-health and health care costs, it is unlikely that these high risks would

display a lower joining preference. Joiners match the expected profile of low risks when examining the results for the utilisation/benefit variables. The probability of joining increases, in particular, with increasing non-essential high-end services and discretionary services, and with decreasing essential benefits.[140]

Contributions and co-payments were, as expected, less able to deliver evidence in regard to adverse selection or risk selection on their own. Higher contributions for joiners are in line with the expectation that insurers charge new members in unregulated markets higher premiums. They do this in order to maximise their profit under the uncertainty of possibly insuring high risks. However, the data subgroup description (in chapter four) showed that joiners (as all other mover groups) have, on average, higher contributions than the control group but also have the lowest claims, yielding a contribution-claim ratio of above one. This result is consistent with the expectation that households with contribution-claim ratios above one are of low risk and are subject to favourable risk selection by insurers. The joining probability increases with decreasing claims and essential benefits, but rising non-essential and discretionary benefits. Thus the rising co-payments can be explained as being generated by low risks, who use non-essential and discretionary services with typically high co-payments. The co-payment variable supports the favourable risk selection hypothesis.

Finally, it was expected that the plan classification variable would be a very powerful variable for providing evidence for selection.[141] The interpretation of these results confirmed the hypothesis that favourable risk selection is operating in the South African medical scheme market. Low risks are selected into the market by the medical scheme plans. Households are much more likely to join low-risk plans as compared to average-risk plans and are less likely to join high-risk plans as compared to average-risk plans. This result, in particular, is important, since most variables already identified joiners as low risks. Given that joiners have a lower expected risk of using health care resources (i.e. lower claims), it can be assumed that favourable risk selection takes place in the market in terms of directing favourable enrolment towards designed low-risk plans.

However, the results do not exactly support the case of a separating equilibrium. As Cutler and Zeckhauser (2000) point out, a separating equilibrium can exist if the

[140] GP benefits also predict joining and joining probability increases with increasing GP benefits. However, it was assumed that, because in most health care systems GPs function as gatekeepers for high-end services used so extensively by low risks, this result indicates that joiners might be, in fact, low risks.

market can efficiently sort high and low risks into either more generous or less generous plans, while charging individuals their expected costs. In the observed South African medical scheme market the insurer seemed to be able to sort high and low risks into either more generous or less generous plans to a certain degree, but they seemed not to be able to charge individuals their expected costs. This point seems to be valid in regard to the contributions in particular. Despite the fact that the risk profile of the joiners is indicating a higher probability of low risks as compared to control group members, the joiners' contributions are significantly higher. This is also true for the co-payment variable, but the later can be better explained together with the benefit structure of joiners.

5.4.3 Model B results interpretation: leavers-control group

First, similar to the socio-economic variables of Model A, age, gender, and number of dependants delivered very good evidence for selection in Model B, while income[142], race, and residency could not be interpreted as evidence for selection processes. The decrease in the likelihood of leaving with age indicates that young, low risks leave the market segment, which at first sight looks like classic adverse selection. However, these low risks can either discontinue insurance coverage altogether (i.e. not only in medical schemes) or resume coverage with other for-profit insurers. It is very likely that this adverse selection was initiated by the cream-skimming efforts of the competing for-profit insurers.

Females have a higher probability of leaving medical scheme plans and the effect of gender on the odds of being a leaver decrease with rising age, particularly for women younger than 40, during their reproductive years. Unfavourable risk selection is likely (in the form of dumping female headed households).[143] Households with no or few dependants were expected to have a lower likelihood of health costs as compared to

[141] However, the odds for above average income households to join a low-risk plan rather than average-risk plan increase were slightly higher than all other odds increases. This supports the favourable risk selection hypothesis that low risks are selected into the market, and specifically low-risk plans.

[142] Below average income households have higher odds of leaving than households with average incomes, matching the expectation that income represents an affordability constraint, but contradicting the expectation that low-income earners have a higher economic incentive to keep insured. It might be that low income leavers seek insurance somewhere else. Then it could be expected that leaving low income households are younger, because as low risks they are likely to be the focus of favourable risk selection by competitors. This could be confirmed: while high income households leave more frequently regardless their age, low income leavers only show odds ratios above one in younger age, when their lower risk allows them to obtain cheaper coverage elsewhere. Households, regardless their income, have increasing odds of leaving high-risk plans. Thus income functions as an affordability constraint for insurance coverage in general and for alternative choices. These results can be interpreted as evidence for strong competition and cream skimming.

[143] It could be argued that for some younger women the likelihood of leaving is so much higher than for men because they get married and obtain insurance through their spouses.

similar households with average dependants number. Nevertheless the former are in fact less likely to leave insurance. Risk-aversion paired with favourable selection might be operating here, where insurers try to keep risk-averse low-risk households in their pool. Since households with an above average number of dependants are of higher risk, their lower probability for leaving can be interpreted as evidence for adverse selection. However, the odds of leaving for the significant interaction dependants*age increase with age, leading to the interpretation that older households with several dependants are more likely to be dumped.[144] Thus while for households with an above average number of dependants the evidence is inconclusive, supporting either adverse selection or unfavourable risk selection, favourable risk selection can be concluded from the result that households with no dependants are less likely to leave insurance coverage.

Claims and benefits variables also provide very good evidence for adverse selection and risk selection in Model B. An increase in households' probability of leaving with increasing claims is here interpreted as unfavourable risk selection. Given the higher likelihood of ill-health and health services usage for people with high claims, it is unlikely that this result reflects a higher leaving preference among high risks. It rather suggests that the insurers dump people with high risk, i.e high claim levels. Probability of leaving increases with decreasing essential and negotiable benefits, but with increasing discretionary benefits. Thus leavers mostly matched the expected profile of low risks when examining the utilisation/benefit variables.

However, the variables of contribution and co-payment only partly supported this interpretation. Increasing contributions support the descriptive results that leavers have a higher mean contribution and a lower calculated contribution-claim ratio (of below one) compared to the control group. This could be interpreted as evidence for unfavourable risk selection by medical scheme plans, dumping high risks. Leaving probability increases with increasing claims and rising discretionary benefits, but decreasing essential benefits. Rising co-payments and increasing claims support the profile of exiting high risks and the hypothesis of unfavourable risk selection. However, the increasing co-payments together with the decreasing essential services but rising discretionary services support the profile of leaving low risks and the hypothesis of adverse selection initiated by external cream-

[144] The increase in leaving probability is, in particular, significant for older households near retirement age with above average dependant numbers. Given these households' higher expected risk and thus difficulties in securing new insurance coverage, frequent leaving seems unlikely to be their preference, unless they exit insurance for natural reasons, i.e. death.

skimming. Thus while the evidence for the co-payment variable cannot interpreted conclusively, a form of risk selection seems to be likely.

Taking the results (from the comparative data subgroup description in chapter four) and the other claim-contribution variables into account, the benefit profile of leavers might match either that of high risks or low risks. Increasing claims and some descriptive results support the profile of high risks and therein the hypothesis of unfavourable risk selection by insurers, that dump high risks out of the South African medical scheme plans. However, increasing discretionary benefits and decreasing essential benefits indicate that a considerable fraction of low risks might be among the leavers. This supports the hypothesis of adverse selection most likely initiated by cream-skimming of medical scheme plans' competitors (i.e. external favourable risk selection). It is very likely that both processes took place simultaneously in the South African medical scheme market.

The results for the plan classification variable were expected to be powerful evidence for either adverse selection or risk selection. They confirmed the previous two-sided interpretation of the estimation results obtained from the leaver-control group model. The result that households are more likely to leave high-risk plans as compared to average-risk plans can be interpreted as confirming the hypothesis of unfavourable risk selection or dumping of high risks, particularly from high-risk plans.[145] However, the results could also confirm the adverse selection hypothesis that low risks leave high risk plans more frequently (possibly initiated by external favourable risk selection), particularly if they can afford coverage somewhere else.[146]

Evidently the data supports the hypothesis of adverse selection *and* risk selection. While there is support for the unfavourable risk selection hypothesis - several variables identified leavers as high risks - some estimation coefficients can only be interpreted in light of adverse selection (probably initiated by external favourable risk selection). Altogether, the South African medical scheme market seems to have been contested and dominated by fierce competition between medical scheme insurers and their for-profit competitors. Thus while the focus was on selecting low risks, high risks were willingly discarded in this process. As with Model A, evidence in support of or against the case of a separating equilibrium can be not concluded.

[145] Also, the lower probability that households leave a low-risk plan, could be attributed to favourable risk selection efforts by the insurers to keep low risks in the medical schemes with designed low-risk plans.
[146] Above average income households tend to leave high-risk plans rather than average-risk plans. This supports the adverse selection hypothesis that low risks exit the market and in particular high-risk plans.

5.4.4 Model C results interpretation: switchers-control group

The socio-economic variables of age, gender and number of dependants delivered less clear evidence for either adverse selection or risk selection in Model C as compared to Models A and B. Income, race, and residency could not be interpreted as evidence for selection processes.

The higher likelihood of the young to switch can be interpreted as favourable risk selection, where insurers try to keep low risks in the pool by switching them to designed low-risk plans. Given the lower risk of ill-health for younger people this makes them a target for favourable risk selection. Females have a higher probability of switching medical scheme plan coverage. Given that females have a higher expected risk of using health care resources than males - particularly in their reproductive years – their greater frequency of switching to high-risk plans at young ages can be interpreted in two ways. First, it might point towards unfavourable risk selection, where female high risks are pressured to switch to more expensive high-risk plans. Second it might be a sign of adverse selection by high-risk females into high-risk plans, in order to secure better and more comprehensive coverage for expected health risks. Thus the interpretation of the gender variable results here is inconclusive.

Households with an above average number of dependants have a significantly lower probability of switching medical scheme plans as compared to households with an average number of dependants. These larger households are expected to be of high risk. Together with the positive sign of the non-significant coefficient for the below average dependant number, this allows the interpretation of the result as evidence for favourable risk selection, where insurers preferably offer low risks opportunities to switch to low-risk plans.

In Model C the claim and benefits variables provided very good evidence for selection. Decreasing claims were interpreted as evidence for favourable risk selection and insurers' efforts to influence low risks to switch into designed low-risk plans. The probability of switching increases with increasing high-end, non-essential and discretionary benefits that were expected to function as luxury goods, while essential benefits decrease with the probability of switching.[147] The results for the variables of contribution and co-payment - on their own less able to deliver evidence in support of adverse selection or risk selection - support the interpretation of the claim and benefit

variables. Decreasing contributions are consistent with the expectation that under favourable risk selection, low risk households switch into less expensive low-risk plans. This result can be interpreted as evidence for favourable risk selection, where insurers induce low risks with low claims, to switch to less expensive low-risk plans. Since the switching probability increases with decreasing claims and essential benefits, but rising discretionary and non-essential benefits, the high co-payments of leavers can be explained as being generated by low risks. Thus the co-payment variable supports the favourable risk selection hypothesis. Altogether the claim and benefit profile of switchers matched that of low risks. This supports the hypothesis of favourable risk selection by the South African medical scheme insurers that selected low risks into low-risk plan options providing a wide range of non-essential, luxury services.

The final variable of plan classification is, as expected, a very powerful variable for adverse selection or risk selection evidence. Switching probability decreases with the probability of a household switching to a plan with an above average risk profile. Again, these results are of great importance, since most variables already identified switchers as low risks. Given that switchers had a lower expected risk of using health care resources (i.e. lower claims), it can be assumed that favourable risk selection takes place in terms of directing households' continuing enrolment towards designed low risk plans.[148] Thus, the interpretation of the risk plan variable confirmed the favourable risk selection hypothesis, that low risks are selected preferentially by South African medical scheme insurers to continue coverage in designed low risk plans.

In Model C the results lend some support to the case of a separating equilibrium. Medical schemes seemed to be able to sort high and low risks into either more generous or less generous plans, and they seemed to be able to charge individuals their expected costs. Matching the fact that the risk profile of the switchers indicates a higher accumulation of low risks in the switcher group as compared to control group, the switchers' contributions and co-payments are significantly lower.

The next section will summarise the results of the statistical models within the context of the research hypotheses.

[147] However, switching probability increases with increasing GP benefits. It was assumed, however, that in most health care systems GPs function as gatekeepers for services typically used by low risks, supporting the interpretation that switchers are low risks.

5.5 Results review in the context of the research hypotheses

The purpose of this section is to review the results of the empirical analysis in the context of this thesis' research hypotheses. The thesis' empirically testable, main research hypothesis sought to explore whether South Africa's unregulated health insurance market encountered a lack of market-wide risk pooling and a lack of risk pooling within the market (between insurers' plan options). Within the case study setting, this main research hypothesis could be formulated into three more precise subhypotheses.

On the next page table 5-10 displays the three empirical hypotheses in the left column and the research hypotheses assessment in regard to the results from the empirical analysis in the right column. The two first research hypotheses assumed, in particular, risk selection activities by the previously regulated but now unregulated medical scheme insurers as a result of competition with other South African for-profit insurers. Hypothesis three assumed adverse selection strategies by high risks. Strong empirical evidence for adverse selection and risk selection under hypotheses one and two was concluded. For example, external favourable risk selection of low risks out of the medical scheme plans leads to a disproportional loss of low risks for the market-wide medical scheme risk pool. Most likely this type of risk selection induces adverse selection. Low risks left medical schemes, and particularly high-risk plans, because they became unattractive, and because competitors offered better choices. However, empirical evidence for adverse selection strategies by high risks was very slim and delivered no support for the third hypothesis that assumed adverse selection. It could be neither concluded that high risks in particular entered (high-risk) medical scheme coverage, nor that high risks switched to high-risk plans in particular.

With table 5-10 and the concluded evidence for adverse selection and risk selection from the South African case study, the research hypotheses can be assessed. Research hypothesis one can be verified. The previously regulated medical scheme market segment is contested, and experienced a disproportionate loss of low risks. It could be shown that external favourable risk selection of low risks was conducted by the for-profit medical scheme plan competitors. They most likely initiated adverse selection in the form of low risks leaving the medical schemes, in particular high-risk plans, by attracting them to low-risk plan options. They most likely initiated adverse selection in the form of low risks

[148] The dramatic decrease in the probability of switching to low-risk plans with rising age supports the favourable selection hypothesis that only expected low risks are given incentives to switch to designed low-risk plans.

leaving the medical schemes, in particularly high-risk plans, by offering them better low-risk plan options.

Table 5-10: Research hypotheses & concluded empirical evidence for the South African case study

Results for empirical hypotheses	Assessment of research hypotheses
Hypothesis 1: Strong evidence for a disproportionate loss of low-risk members was shown by testing the leavers and the control group for differences in their risk profiles. Medical schemes lost disproportionately low risks; it was observed that: A. Assuming risk selection by (external for-profit insurers), leavers were, for example, **younger** and had **lower claims** than control group members; and (initiating) B. Assuming adverse selection where low risks exit medical schemes, leavers were, for example, **younger**, had **lower claims**, and had more likely a **high-risk plan membership** than control group members.	Hypothesis 1: Verification of competition for low risks between medical schemes and for-profit insurers. The previously regulated medical scheme market segment is contested, and experiences a disproportionate loss of low risks due to: A. Risk selection by (external) for-profit insurers for low risks in the medical scheme market segment took place, and (initiating) B. Adverse selection with low risks exiting medical schemes, and, in particular, high-risk plan options Medical scheme coverage probably became comparably expensive and more suitable options were offered elsewhere.
Hypothesis 2: Strong evidence for med. schemes' risk selection strategies was shown by testing joiners/switchers/leavers and the control group for differences in their risk profiles: A. Assuming risk selection that discourages the entry of high risks, joiners were, for example, **younger** and had **lower claims** than control group members; B. Assuming risk selection that attracts the entry of low risks, joiners were, for example, **younger**, had **lower claims**, and had more likely a **low-risk plan membership** than control group members; C. Assuming risk selection that prevents the exit of low risks, switchers were, for example, **younger**, had **lower claims**, and had more likely a **low-risk plan membership** than control group members; *D. No evidence for risk selection in the form of pressuring high risks to switch; with switchers having, for example, older, had higher claims, and had more likely a high-risk plan membership than control group members; and* E. Assuming risk selection by discouraging high risks to continue their medical scheme coverage, leavers were, for example, older, had **higher claims**, and had more likely a **high-risk plan membership** than control group members.	Hypothesis 2: Verification of competition, and verification of several hypothesised strategies applied by medical schemes to protect their scheme and plan risk pools: A. Risk selection was concluded in form of discouraging high-risk insurance applicants from entering medical schemes; B. Risk selection was observed in form of low risks being attracted to enter low-risk plans; C. Risk selection took place, with low risks switching into attractive low-risk plan options, preventing them from leaving; D. *Risk selection in the form of pressuring high risks to switch into high-risk plan options could not be concluded; and* E. Risk selection by discouraging high risks to continue their medical scheme coverage is likely, particularly if leavers belong to high-risk plans.
Hypothesis 3: Insignificant evidence for adverse selection strategies of high risks by testing joiners/switchers and the control group for differences in their risk profiles: *A. Under adverse selection joiners are, for example, older, have higher claims, and more likely have a high-risk plan membership than control group members; and* *B. Under adverse selection switchers are, for example, older, have higher claims, and more likely have a high-risk plan membership than control group members.*	*Hypothesis 3: No verification of the hypothesised adverse selection strategies of high risks, either seeking medical scheme coverage or already insured here:* *A. Adverse selection by choosing to enter medical scheme plan coverage, particularly expensive and comprehensive high-risk plan options; and* *B. Adverse selection by choosing to switch into medical scheme plans with expensive and comprehensive coverage, typically high-risk plan options.*

Research hypothesis two can be also verified for most part. Medical scheme plans applied various risk selection strategies in order to protect their risk pools. They discouraged high risks from entering medical scheme plan coverage and dumped already insured high risks. However, evidence could be not confirmed for unfavourable risk selection in the form of pressuring high risks to switch to high-risk plans. Medical scheme plans favourably selected low risks, by offering self-selecting low-risk plans for new low risk applicants or offering already insured low risks the opportunity to switch to low-risk plan options. Research hypothesis three cannot be verified. No or insufficient evidence was found for high risks either seeking to enter medical scheme plans, in particular high-risk plan options, or insured high risks trying to switch to high-risk plan options. It could be argued that the dominant risk selection strategies diminished adverse selection processes.

In conclusion of this subhypotheses assessment, the main research hypothesis could be verified. South Africa's unregulated medical schemes encountered a lack of risk pooling. The results indicated that first, the whole observed medical scheme market segment experienced a lack of risk pooling due to risk selection. Risk selection was an expected outcome, because in such unregulated health insurance markets (as in the case of the South African medical scheme market) it is legal to risk select. If medical scheme plans *and* other for-profit insurers can now practice risk selection, it is not surprising that there is competition for low risks. Since for-profit insurers could begin risk selecting much earlier, they obviously had a competitive advantage and were able to extract a considerable number of low risks from the medical scheme market. The results of the empirical analysis - that many low risks leave medical scheme plans - can be best read as a direct consequence of competition between the medical scheme plans and for-profit insurers.

Second, within the observed medical scheme market segment the ever more competitive environment led to a lack of risk pooling. If some plans were better at identifying and dumping high risks or identifying and attracting low risks they had a competitive advantage compared to the rest of the medical scheme plans. They would be more successful in pooling low risks as a result of better risk selection. Similarly, other plans that did not implement risk selection procedures would very likely become high-risk pools over time. The results of the empirical analysis could confirm that a distinction between low- and high-risk plan options, which pool different risks, is possible. The results also confirmed that over the four years of observation high-risk plan options

pooled high risks and low-risk plan options pooled low risks. However, the most interesting result here was that the drive for this risk pool disparity between medical scheme plans was due to the plans' own diverging risk selection strategies. Thus here too the results can be interpreted as a direct consequence of higher competition between different medical scheme plans.

Last but not least, adverse selection did not contribute to the lack in risk pooling. In light of the overwhelming evidence for risk selection this is, however, a rather expected result since risk selection combats adverse selection.

The result of the three models could not be very well interpreted within the framework of a separating equilibrium. Although to a certain degree South African medical schemes seemed to be able to sort high and low risks into either more generous or less generous plans, they seemed to be not able to charge individuals their expected costs. For example, despite the fact that the risk profile of the joiners indicated a higher accumulation of low risks in the switcher group as compared to control group, the joiners' contributions and co-paymets were significantly higher. Only in the switcher-control model did the low-risk profile of the switchers match the switchers' significantly lower contributions and co-payments. This might indicate that plans more efficiently separated existing members, but not new medical scheme applicants. However, in the data sample used for this thesis' analysis switchers only represent a very small fraction of the observed medical scheme membership - 3.7% of all households. Also, the data analysis did not focus on the analysis of trends over time for membership or risk pooling, partly because the data observation period was quite short. No definite test or assessment for the existence of a separating equilibrium can be concluded due to these limitations to data and data analysis design. However, if the results are interpreted in light of this hypothesis, the formation of a separating equilibrium for the South African medical scheme market seems unlikely.

All results of this thesis' analysis have now been presented and reviewed in the context of the research hypotheses.

5.6 Summary

This chapter fulfilled this thesis' research objective by presenting the results of the empirical investigation into whether unregulated health insurance markets, such as South

Africa's medical scheme market, experience adverse selection and/or risk selection. The presented results could be interpreted as empirical evidence for selection. Since the results of the analysis yielded evidence for selection processes it could be concluded that the empirical method, which utilised available quantitative techniques and was applied to easily accessible data, fulfilled its purpose successfully. This means that both research aims - first, deriving of an easily applicable method for empirical investigations and second, testing of this method on a case study data set - were effectively accomplished.

The first part of this chapter presented the estimation results for the three logit models - Model A, which compared medical scheme plan joiners and the control group, Model B, which compared leavers and the control group, and Model C, which compared switchers and the control group.

The second part described the results for the model fit assessment. However, since diagnostic statistics for the overall fit of a logistic regression model with correlated data, such as used here, have not yet been developed, the literature suggested approximating the model fit assessment with standard methods developed for uncorrelated data models. The thorough model fit assessment of the proxy models was evaluated as being good. Based on this, the conclusion was drawn that the actual marginal models used, which are more appropriate for the correlated data, were fitted very well.

The fourth part presented the interpretations of the model estimation results and concluded evidence for selection in South Africa's unregulated medical scheme market between 1995-1998. Three main inferences can be made, based on the interpretation of the statistical model results. First, overwhelming evidence for favourable selection of low risks into to the medical scheme plans could be derived from the joiner-control group Model A. Support for favourable risk selection practised by medical scheme plans arose also from the switcher-control group Model C. Second, evidence for unfavourable selection by medical scheme plans in the form of dumping could be concluded from the leaver-control group Model B. Third, evidence for adverse selection was not found.

Finally part three of this chapter reviewed the analysis results in the context of the research hypotheses. Research hypotheses one and two were verified. First, competition between the medical scheme plans and for-profit insurers caused risk selection, breaking established risk pooling arrangements. Second, medical scheme plans applied various risk selection strategies, protecting their risk pools, with some plans pooling increasingly high risks and other plans pooling increasingly low risks. Research hypothesis three was not verified due to insufficient evidence for adverse selection, specifically the absence of the

typically expected adverse selection cycle. Thus this thesis' main research hypothesis was verified: South Africa's unregulated health insurance markets encountered a lack of market-wide risk pooling and a lack of risk pooling within the market.

The empirical analysis of this thesis is concluded. The next chapter will present this thesis' discussion including policy recommendations for the South African case study.

Chapter 6

Discussion: thesis contributions, limitations & policy relevance

6.1 Introduction

The purpose of this chapter is to discuss this thesis' contributions, selected limitations and policy relevance, which will enable the assessment of the research question. The thesis applied a method that yielded testable hypotheses for empirical investigations into adverse selection and risk selection to an empirical analysis of an unregulated health insurance market. This is the major contribution of this thesis. In doing so, a number of limitations have been identified. From the results of the empirical analysis for the South African case study it is possible to derive the thesis' policy relevance. First, it is the aim of this chapter to discuss the thesis' main contributions. Second, limitations of this thesis, mostly related to the data and empirical analysis, will be discussed. Finally, the thesis' policy relevance will be assessed and the originally posed research question answered.

Part two of this chapter will discuss the group approach and the empirical analysis as this thesis' main contribution. The generalisability of the developed method will be assessed, based on the data requirements of either individual-based cross-sectional or panel data. It will be shown that the group approach is generalisable to different country settings and various insurance arrangements. In particular, it can be assumed that it is easily applicable to low and middle-income country-settings, because of the limited data requirements. This was demonstrated by applying it to the case study data from South Africa.

The third part of this chapter will mainly discuss limitations of the data and of the empirical analysis performed in this thesis. However, the assumption of rationality and

the inference of motivation from observed outcomes of behaviour are the fundamental limitation of this research, aside from the technical limitations of data and analysis. First, these two issues will be shortly discussed before turning in more detail to the data limitations from the South African case study. Data subgroup and plan classification and issues related to the formulated expectations of the risk-profile variables will be considered. Second, a discussion on limits to interpreting the results of the empirical analysis that had to be considered in order to derive meaningful interpretations will be included. Finally, limitations of the thesis' model-fit assessment will be discussed. The problem of absent diagnostic statistics for the fit-assessment of logistic regression models for correlated data will be discussed as well as the strengths and weaknesses of the different summary measures of goodness-of-fit and logistic regression diagnostics, also presented previously.

The fourth part will discuss the thesis' policy relevance. First, the general political relevance will be considered. It will be shown that the results support the research hypotheses, because causality between the observed results and concluded risk selection can be assumed. Possible explanations for this thesis' most interesting result, the overwhelming evidence for risk selection but absence of evidence for adverse selection, will be discussed. The results will also be assessed in light of this thesis' research question - whether of evidence of risk selection can be interpreted as increasing inequity. Second, the specific policy relevance of the thesis' empirical results for the South African case study will be discussed. Two strategies for political action proposed and implemented in order combat increasing inequity in South Africa's health insurance arrangements will be shortly discussed. Some further policy recommendations derived from this thesis' results will be formulated.

6.2 Discussion of the thesis' contributions

The purpose of this section is to discuss the main contribution of this thesis, the application of the group method to an empirical investigation that seeks to identify adverse selection and risk selection in easily accessible data. The method emphasises the group dimension of these selection processes, which was underrepresented in previous research. It was the objective to highlight this particular dimension of selection processes through which it becomes feasible to examine these complex processes empirically. The

aim was to break complex selection processes down into empirical problems that are easy to explore analytically.

The literature acknowledges that, as a result of adverse selection and risk selection, health insurance risk pools change. Theoretical concepts of both selection processes, which are used in the literature, explain how people's motivation to select into specific health insurance plans leads to different risk pools. It was therefore intitutive to look at individual (or household) selection behaviour, which consists of joining, leaving and switching between health insurance plans. The risk pool of health insurance plans can only change in two ways: either because the risk profiles of the same individuals within the pool change, or because individuals who move into, out of, or between risk pools contribute to risk profile changes. Thus during selection processes people with certain risk profiles self-select or become subject to insurers' selection into or out of certain risk pools. Previous empirical investigations, like the multiple plan choice approach, compared individual characteristics of people who choose insurance plans of different generosity levels. Some of these studies divided the insured population into the subgroups of possible carriers of the selection process - people who move into, out of, and between insurance plans - and compared their characteristics to the rest of the insured population.

This thesis contributes to the empirical literature on adverse selection and risk selection with an application of the group method and the execution of an empirical analysis of South Africa's unregulated health insurance market. Adverse selection and risk selection in health insurance markets are of academic and political concern. There are several studies on these issues for high-income countries but there are rarely any for low- and middle-income countries. This empirical analysis adds to these few existing studies, thereby increasing the understanding of selection processes in health insurance markets in a low- and middle-income country context. However, in conducting the thesis' empirical analysis a number of limitations have been identified. Most were related to the real world data, which is usually flawed or even non-existent. Thus data represents the main limitation of the group approach, because data is often imperfect. The next section will discuss some of the limitations that were faced during the empirical analysis, including data limitations.

6.3 Discussion of selected thesis limitations

The purpose of this section is to discuss selected limitations of this thesis, particularly in regard to the data used and the empirical analysis performed. Nevertheless, in addition to these technical limitations, the two fundamental limitations of research of this kind: (1) the assumption of rationality and (2) the inference of motivation from observed outcomes of behaviour will be discussed first. This thesis employed an economic approach and accepted the general model of rational economic behaviour as its fundamental underlying assumption. The empirical analysis, its results and interpretations presented here are based on the fact that only consequences or outcomes of behaviour are observed, not motivations.

The assumption of rationality can be a powerful assumption for explaining behavioural outcomes. However, economists do not literally believe that everyone always behaves perfectly rationally. For example, the model of economic behaviour is not based on perfect information. Although an individual's decisions and actions are mainly driven by information, imperfect information does not mean that individuals are acting irrationally or unpredictably. The assumption here is rather that the accumulation of information is only possible under certain cost constraints, and that rational individuals will only collect additional information as long as their expected marginal utility is greater or equal to the marginal costs of doing so. However, this information equilibrium is different for different participants in the market, and the phenomenon of asymmetric information may occur. In a health insurance market this might lead to selection processes, which not only destroy efficient competitive market outcomes and potentially lead to market failure, but also constitute grave inequities that might call for political action. This means that the rationality concept is quite precise and can be prescriptive. Thus this thesis' economic approach complements political science, sociology, psychology and other disciplines.

The fact that only consequences or outcomes of behaviour, and not motivations, are observed is a limitation that all disciplines face. The thesis' economic approach, as with other scientific approaches, is built on interplay between theory and observable data. Models are built to help explain the data and data are used to help refine models. However, data and analytical limitations can influence the inference from observed outcomes of behaviour to its motivation.

This section will proceed by first discussing specific data issues related to the South African case study. Second, issues related to the empirical analysis will be

discussed, namely limits on the interpretation of the results. Finally, the model fit assessment will be discussed, focusing on the limits of the different presented measures of goodness-of-fit and logistic regression diagnostics.

6.3.1 Limitations of the South African data

As with all retrospective studies, data here was limited to what has been recorded. Also, the data was not designed for assessing risk pooling and selection processes. Some data limitations were previously discussed in the data section in chapter three. This concerned, for example, the possible selection bias of medical scheme plans. The data set provided by the medical scheme plan management firm contained a given set of medical scheme plans, and, although it was assumed that this sample is representative, this might not be the case, which could have biased the results. Another important weakness is the lack of control for environmental or organisational factors that might affect plans' risk pooling. Plan size and plan management variables could be not considered in this study. Omitting these factors may have biased results.

One important limitation of the data was its short time horizon. Unfortunately the data was only available for four years of observation. It might have been useful to present alternative model results (for example logistic regressions by individual years). Originally it was planned to run alternative regressions by individual years. However, the extensive data examination prior to the data analysis did not suggest more exploitable yearly variations in the data between 1994-1998. Therefore the plan of yearly or other time-related split-sample analysis was abandoned. The limited time horizon of the data is also related to the limited ability to test for a separating equilibrium. As expressed in chapter five, the data analysis did not focus on the analysis of trends over time for membership or risk pooling, because the data observation period was quite short. A longer time-dimension of the data and the analysis of time-trends might have better lent itself to the assessment of a separating equilibrium hypothesis.

The discussion of the limitations of data used in this thesis' analysis will focus on limitations related, first, to data subgroup and plan classification and second, to the variables of the risk profile.

First, the data subgroup classification divided insured households into: (1) households who newly enrolled into medical scheme plans during the observation period, i.e. the joiners, (2) households who disenrolled from medical scheme plans, i.e. the leavers,

(3) households who switched between plans, i.e. the switchers, and (4) households who remained constantly insured in one plan, i.e. the control group. One procedural problem here was that the length of the observation period for households varies, since some joined, left, or switched in the first or second year or some only moved during the fourth year. This means that, on average, joiners, leavers, and switchers have shorter observation periods than the control group (which observation period is by definition four years). However, the statistical estimation method that was applied did consider this type of censored or incomplete data. Another problem arose when households joined or switched just before the year ended or left at the beginning of a year. For example, households who entered at the beginning of a year have observations over several months as compared to households who just joined a month before the end of a year. However, calculating all time-invariant variables other than age, such as contributions, co-payments, claims and benefits as per member and membership month for each household solved that problem.

Nevertheless, other problems remained. While households that just joined or switched the year before the observation period started, in 1995, or left after the data time series ended in 1998, were counted into the control group, households that were actually observed to join in the beginning of 1995 were counted into the joiners group. Also, it could be not considered that there might be households that moved more than once, for example, households that joined or switched and then left. Each household received a new ID upon joining and this ID ceased to exist upon leaving. Thus the same household that left one medical scheme could have joined another medical scheme under a new ID during 1995-1998. This means that it remained unobservable where joiners or switchers came from or to where leavers went. For individuals who joined a medical scheme, it is not clear whether they came from another medical scheme inside or outside the sample, from the external for-profit insurers or from the uninsured population.

For example, leavers might have left insurance coverage completely, they may have joined another medical scheme within or outside the sample, or they may have joined for-profit insurance. Indeed it is possible that they left insurance with the schemes because of a life-terminating event. This has a particular implication for the hypothesis that under risk selection older people are more frequently dumped from insurance. Older households might terminate medical scheme coverage more frequently than other households because of the death of the principal member. For the switchers it was only possible to observe households that joined into a new plan within a scheme or left a certain plan for another within the same scheme but not from which plan they came or

which new plan they joined. Again, this is due to the fact that the ID number of each individual change with each membership move because it is connected to the particular plan they join, leave, or switch to. Also, it was pointed out that most switcher observations were for joining switchers (90%).

Another data subgroup classification concerned the plan classification into different risk pools. This classification of plans into low-, average-, and high-risk plans depended on: (1) the concluded ability of variables to function as reliable risk-adjusters, (2) the definition of the mean intervals that were the basis for this classification, and (3) the judgement of whether the means displayed a consistent pattern indicating an accumulation of either low risks or high risks. This plan classification process had its limitations, i.e. it was to some extent subjective, but it was not arbitrary, since it followed a classification algorithm spelled out in chapter 4. The ability of variables to function as reliable risk-adjusters was based on findings in the risk adjustment literature. In order to support the plan classification algorithm, based on the definition of the mean intervals, a sensitivity analysis was performed, which indicated that the plan classification was robust even when the mean interval increased or decreased by 10%. Probably the last part of the classification algorithm, the judgement of whether the means displayed consistent pattern indicating an accumulation of either low risks or high risks carried most subjectivism.

For example, plan 17 in table A5-2 in Appendix 5 was classified as an average-risk plan, despite having for five variables above average variable means. Plan 17 had higher than average discretionary benefits, but only average total claims. Also, the mean age was below the overall average. Thus the pattern was not consistent with high-risk accumulation. On the contrary one might have suspected that this plan attracted young low risk with generous discretionary services. However, plan 17 had above average specialist benefits and the average total claims were near the upper mean interval border. Taking all these observations together it was decided that there was no consistent pattern that could be derived from the overall plan means. The means pattern for plan 17 was inconclusive and that is why it was not included into the group of low or high-risk plans, but classified as an average-risk plan.

Second, there are limitations regarding the variables used as risk indicators for the households' risk profiles. For example, the age variable could be loaded with several contrasting effects. On one hand, it could be expected that older households leave medical scheme plans due to death, but this remains unrecognised in the data. On the other hand, young households are expected to enter medical schemes more often, because they join

the workforce. Effects like these can diminish the usefulness of the variable age for identifying adverse selection or risk selection effects. Similar problems might arise for other variables. For instance, the gender of the head of household is strongly influenced by the workforce composition and compensation practices that allow coverage of females with their spouses and vice versa. However, these limitations were considered in the formulation of the variable expectations under adverse selection and risk selection. Most variables that were affected by this problem were only included as control variables.

Another possible limitation includes the coding of the income variables. First, the original income variable provided no absolute Rand values, but instead had coding for approximate income group categories between zero and 999. This coding was the only income indicator provided by the medical schemes holding company, which assigned these categories. Confidentiality reasons were given as an explanation for the lack of more precise income data. These income categories only provided a relative income value, with zero being the lowest income category and 999 the highest. For the analysis the income variable was converted into a categorical variable with only three categories, average, below and above average. It was hoped that this would provide more easily interpretable results, with the reference category defined as the average income group in the empirical analysis. This coding produced uneven income groups. The average income group was defined as households of the income categories 500-599. The below average income group was defined for the income categories 0-499 and the above average income group for the income categories 600-999. In retrospect it would have been better if the average income group would have been coded wider. However, the income variable was not a key variable for the purpose of this analysis, but rather functioned as an important control variable. It is unlikely that the uneven income groups resulting from this coding had an influence on other analysis results. Also, the results for the income variable were not used for the deduction of possible adverse selection or risk selection.

6.3.2 Limits on the interpretation of the analysis results

This discussion focuses on the empirical analysis and the limits to interpreting its results, in particular, the limited ability to interpret the results within a framework of a separating equilibrium, issues regarding the nature of relationships in statistical models, and other considerations in order to obtain meaningful interpretations results.

In chapter 5 it was expressed that the result of the three models could be not very well interpreted within the framework of a separating equilibrium. Specifically, only the switcher-control model supported the separating equilibrium hypothesis. However, switchers only represented a very small fraction of the observed medical scheme membership (3.7% of all households) in this thesis' data sample. Also, due to the previously discussed short data observation period, the data analysis could not focus on time trends for membership or risk pooling. Thus data with longer time horizons might better lend itself to the assessment of a separating equilibrium hypothesis.

After estimating coefficients it is of concern to assess the significance of the variables in the model (Hosmer and Lemeshow 2000). This concerns the statistical hypothesis, which asks whether the independent variables in the model are significantly related to the dependent variable. Regression analysis evaluates the contribution of each variable to the model by testing for its statistical significance and then examines the significance of its effect on the dependent variable (Menard 2001). In logistic regression the method for evaluating the statistical significance of the contribution of an independent variable to the explanation of a dependent variable is either the likelihood ratio test or the Wald statistic.[149] The regression coefficient is then examined to determine whether the change in the dependent variable associated with a given amount of change in the dependent variable is large enough to be concerned about (Menard 2001). The point of importance here is that this concerns an association, and does not imply a causal relationship. Relationships that are described by statistical models are predictive but may or may not be causal (Menard 2001).

For the estimated models in this thesis, statistical significance of the contribution of most independent variables to the explanation of variations in the dependant variable was found. As expected, several independent variables acted as good predictors for a household being a joiner, leaver or switcher, rather than a control group member. However, some variables might have been effects as easily as causes (for instance having belonged to a high-risk plan). Thus the group membership predictors could be interpreted as effects or causes for a household's likelihood of belonging to one of the data subgroups. However, these results are best interpreted without implying a causal relationship. In a statistical sense all that can be said is that these variables are associated with a certain group membership.

[149] The guiding principle here is to compare observed values of the response variable to predicted values obtained from models with and without the variable in question (Hosmer and Lemeshow 2000).

Furthermore, the interpretation of results always concerns a second issue besides determining the functional relationship between the dependent and independent variables, namely, defining the appropriate units of change for the independent variable (Hosmer and Lemeshow 2000). The latter is important because the estimated coefficient for the independent variable represents the slope (or rate of change) of a function of the dependent variable per unit change in the independent variable. In logistic regression the slope coefficient represents the change in the logit corresponding to a change of one unit in the dependent variable (Hosmer and Lemeshow 2000). In order to derive correct interpretations the differences between two logits have to be meaningful. For the models used in the empirical analysis of this thesis, the estimated coefficients for the independent variables were of main interest, while the intercepts were of lesser interest. Some variables were dichotomous (e.g. gender), others polychotomous (e.g. race), and several were continuous. Some issues related to the most widely used measure of association, the odds ratios, are discussed here.

First, the odds ratio for a variable is obtained by exponentiating its estimated coefficient. The odds ratio is then used as a measure of association. In the case of a dichotomous variable it approximates the likelihood of an outcome - how much more or less likely it is - to be present among those with $x=1$ than among those with $x=0$. However, two important limits should be kept in mind when interpreting the odds ratios. Number one is that often the interpretation of the odds ratio is performed in terms of relative risk[150], based on the argument that it approximates the relative risk in many instances (Hosmer and Lemeshow 2000). But this is only true under certain fairly restrictive conditions (of very small odds). In general, using the odds ratio to represent a risk ratio will overstate the strength of the relationship (Menard 2001). Thus an odds ratio of 2 for females that join insurance compared to males does not necessarily mean that the *risk* of joining is about two times as high for females as for males.[151] Limit number two is that although the odds ratio is the parameter of interest in most logistic regressions (because of its easy interpretation) its estimate has a skewed distribution. However, for large sample sizes, such as used in this thesis, the distribution of the odds ratio estimate is normal (Hosmer and Lemeshow 2000).[152]

[150] The odds ratio is a ratio of two odds while the risk ratio is a ratio of two probabilities.
[151] This approximation can often be misleading and should be avoided. Besides, if needed, the relative risk can be calculated separately (see Menard 2001).
[152] Since in most cases the sample size requirements are too big, inferences are usually based on sampling distribution of the odds ratio logit that follows a normal distribution for even small sample sizes.

Second, most computer packages automatically provide odds ratio estimates (and their confidence interval) based on the exponentiation of each coefficient. However, these estimates are only of interest for dichotomous variables coded zero or one with no interactions. For all other variable types (polychotomous and continuous variables), with or without interactions, specific methods have to be applied to calculate odds ratios of interest that provide meaningful interpretations.[153] These methods were applied in this thesis' chapter five when calculating odds ratios for the results and subsequently interpreting them.[154] Here some issues regarding the interpretation of odds ratios for continuous variables are pointed out as examples.

The interpretation of the estimated coefficient for continuous variables depends on how that variable entered the model and the units of the variable. The slope coefficient gives the change in the log odds for an increase of one unit in x, which is of little interest in most settings (Hosmer and Lemeshow 2000). For example, in this thesis's models a one year increase in age or one Rand increase in claims is too small for a meaningful interpretation. Similarly, a change of 50 or 100 Rand seems too large. A ten year change in age to be most meaningful seemed in the context of the analysis, because it represented an age difference that can be easily understood and interpreted. Hence, odds ratios (and confidence intervals) were calculated for the continuous scale covariates for a change of ten units in the covariates (10 Rand for all claims, contributions, and benefits and 10 years for age).

However, interpretations derived from these estimated odds ratios are based on the assumption that the logit is linear in the continuous covariate.[155] In order to be able to derive interpretations valid over the whole range of the continuous variable, the linearity in the logit for each continuous variable had to be tested beforehand.[156] Further, if there is interaction between two variables, the estimates of the odds ratio for the first variable depends on the value of the second variable with which it is interacting. In cases where interactions are present the correct model-based estimates for the odds ratios can be obtained by calculating the difference between the two logits and then exponentiating the

[153] For an excellent description the interested reader is referred to Hosmer and Lemeshow (2000).
[154] The mathematical equations for their calculation can be obtained from Hosmer and Lemeshow (2000).
[155] For example, it is expected that the additional likelihood of being a joiner for a 20 year-old compared to a 30 year-old is similar to the additional likelihood for a 50 year-old compared to a 60 year-old.
[156] For example, it was shown that age can be treated as linear in the logit, making the interpretation of the odds ratio for a 10 year change in age valid over all ages.

value.[157] For the results interpretation in chapter five, relevant interaction between a risk factor and another variable was calculated and incorporated into the interpretation.

6.3.3 Limits of the model fit assessment

The main limitation of the model fit assessment was pointed out in chapter five. Due to the lack of diagnostic statistics for the overall fit of logistic regression models for correlated data, the model checking was performed with proxy-models for uncorrelated data. These methods, developed for non-correlated data, limited the model fit assessment in terms of its interpretational value, because they really assess the fit of a different model (and not the one from which the results were obtained). However, the main focus of the discussion here is on the limits of the measures of goodness-of-fit and logistic regression diagnostics.

A. Summary measures of goodness-of-fit

Summary measures of goodness-of-fit are functions of residuals defined as the difference between the observed and fitted value (y-$\bar{y}$). The difference between the observed and the fitted value can be measured in several ways in logistic regressions. However, fitted values are calculated for each covariate pattern and depend on the estimated probability of the covariate pattern (Hosmer and Lemeshow 2000).

There is no widely accepted direct analogue to the R^2 of linear regression models. This is because an R^2 measure seeks to make a statement about the "percentage of the variance explained", but the variance of a dichotomous/categorical dependent variable depends on the frequency of distribution of that variable. For a dichotomous dependent variable, variance is at a maximum for a 50-50 split and the more lopsided the split, the lower the variance. This means that R-squared measures for logistic regressions with differing marginal distributions of their respective dependent variables cannot be compared directly. Comparison of logistic R-squared measures with R^2 from linear regressions is also problematic. Also, the pseudo R^2 measure does not strictly assess goodness-of-fit of a model but compares the predicted values from the fitted model to a model with no data or an intercept-only model (Hosmer and Lemeshow 2000). Thus it is more effective to attempt to measure strength of association. Nonetheless, a logistic R-squared measure is available in most software packages as Pseudo-R-Square.

[157] For the algebraic details see Hosmer and Lemeshow (2000).

Two other goodness-of-fit tests for logistic regression models offered in most software packages are the Pearson goodness-of-fit chi-square statistic and the Hosmer-Lemeshow goodness-of-fit test.[158] The Pearson chi-square goodness-of-fit test is a test of the observed against the expected number of responses using cells defined by the covariate patterns (STATA 1999). However, if the number of covariate pattern is close to the number of observations the Pearson chi-square test is not appropriate and the Hosmer-Lemeshow goodness-of-fit chi-square test might be a better choice (Hosmer and Lemeshow 2000).[159] However, in cases where these tests indicate an acceptable model fit this does not mean that the model necessarily explains much of the variance in the dependent, only that what it does explain is significant. As with other tests, as the sample size gets larger, the test's power to detect differences from the null hypothesis improves. Generally, these measures of fit only provide a rough index of whether a model is adequate. There is no convincing evidence that selecting a model that maximises the value of a given measure leads to a model that is optimal in any sense other than having a higher or lower value of that measure (Long and Freese 2001).

Often individual components of the summary statistic are examined. For example, classification tables are obtained by cross-classifying the outcome variable with a dichotomous variable whose values are derived from the estimated logistic probabilities. A cut-point is defined (most commonly 0.5) in order to compare each estimated probability to it. Estimated probabilities are thus used to predict group membership and it is assumed that if the model predicts accurately then there is evidence that the model fits. Unfortunately, this is not always the case. For example, a logistic model may fit, but still have a poor classification (Hosmer and Lemeshow 2000). Also, classification always favours classification into the larger group, independent of a model's fit. This is the case because classification is sensitive to the relative size of the two groups. Thus classification tables are not very useful measures of model performance. Another reason for the limited function of classification tables as measures of model performance is that they depend heavily on the distribution of the probabilities in the sample. Sensitivity and specificity are declining with more heterogeneous populations and tend to be most accurate for polarised populations. Differences in the sensitivity and specificity of two compared

[158] For a more detailed mathematical description of the pseudo R-squared measure and these two goodness-of-fit test statistics see Hosmer and Lemeshow (2000) and STATA Manual (1999:Vol2).

[159] The Hosmer-Lemeshow goodness-of-fit test divides subjects into deciles based on predicted probabilities and computes a chi-square from observed and expected frequencies. Then a probability (p) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model.

models may be therefore due to the different population-mix and not their performance (Hosmer and Lemeshow 2000).

It is important to realise that models with summary measures of goodness-of-fit that indicate a poor fit may still discriminate well. In the end the assessment of model performance is in equal parts determined by both the summary measures of goodness-of-fit and discrimination. Overall goodness-of-fit measures and classification tables are often used as useful supplements for the assessment of fit, because of their advantageous ability to provide single numbers that summarise a considerable amount of information. However, given the model, the context, and the research objective, some measures perform better, some worse.

B. Logistic regression diagnostics

Several logistic regression diagnostics were performed in order to see if a fit was supported over the entire covariate pattern. Logistic regression diagnostic techniques can easily identify covariate patterns that are poorly fit and those that have a great deal of influence on the values of estimated parameters. However, they rely purely on the visual assessment of plots (for instance $\Delta\chi^2$ or $\Delta\hat{\beta}$ versus the estimated logistic probability). This is because the distribution of the diagnostics under the hypothesis that the model fits is only known within specific limits (Hosmer and Lemeshow 2000). All diagnostics in this thesis were evaluated by covariate patterns. [160] Hence, approximations of their distributions based on the normal distribution (under binomial errors) depend on the number of subjects with that pattern (Hosmer and Lemeshow 2000). However, Models A, B, and C contain some continuous variables and the number of covariate patterns and number of observations is nearly equal. In this case these statistics cannot be relied upon (Hosmer and Lemeshow 2000). Thus the diagnostics statistics are hard to assess in terms of what sustains a large or small value. Hosmer and Lemeshow (2000) argue that in the end this assessment has to be a judgement based on each particular data set (as it was in this thesis).

Logistic regression diagnostics results are optionally used to delete observations with extreme covariate patterns in order to improve the fit of the models. However, in this thesis all goodness-of-fit measures and logistic regression diagnostics were applied to

[160] STATA™ calculates all diagnostic statistics (and residuals) in terms of covariate patterns (m), not observations (n), i.e. all observations with the same covariate pattern are given the same diagnostic statistic (STATA 1999). Hosmer and Lemeshow (2000) argue that this m-asymptotic statistics is more appropriate than n-asymptotic statistics.

simple logistic models. The later only functioned as proxies for the actual correlated data models in order to make up for the lacking fit assessment tools for models with correlated data. For these reasons the diagnostic was only used for visual assessment of the fit of the proxy models, assuming uncorrelated data.

The next part will discuss the policy relevance of the analysis results.

6.4 Discussion of policy relevance

The purpose of this section is to discuss the policy relevance of this thesis' results. This section will first discuss the general policy relevance of the results and how these results address the research questions that motivated this thesis. This concerns, in particular, issues related to the causality and the generalisability of the thesis' results. Second, the specific policy relevance of this thesis' empirical results for the South African case study will be discussed by focusing on two strategies for political action that have been proposed.

6.4.1 General policy relevance of the results

In the last chapter the interpretation of the empirical results led to the conclusion that selection processes existed in South Africa's unregulated medical scheme market. The first question that needs to be discussed here is if causality between the results and their interpretation as evidence for the two selection processes can be assumed. Thus it needs to be discussed whether or not the results from the empirical analysis indeed support the research hypotheses. The second issue would be to assess the results in light of the research question. Can the analysis results for adverse selection or risk selection evidence be interpreted in terms of increasing inefficiency or inequity in South Africa's unregulated medical scheme market? The answer to these questions determines the general policy relevance of the results from the South African case study, in particular, their generalisability to similar settings in other countries.

A. Causality between results and their interpretation as evidence for selection processes

The fifth chapter assessed this thesis' three research hypotheses and concluded that there is primarily evidence for risk selection. Hypotheses one and two assumed that due to competition and risk selection, the South African unregulated medical scheme market would encounter unstable risk pooling. The observed results confirmed anecdotal reports

of aggressive underwriting and thus risk selection practices by the medical schemes' for-profit competitors. As a result households that indicated a low risk profile left medical scheme plans, and, in particular, high-risk plans. Although without a context these results look like classical adverse selection, it could be shown that for this thesis' data it was external favourable risk selection. As another result of competition medical scheme plans dumped high risks and discouraged their entry. Dumping was demonstrated by the higher likelihood of households with high-risk profiles to leave high-risk plans. Discouraging high-risk entry was demonstrated by showing that households with high-risk profiles were less likely to join medical scheme plans and less likely to remain in medical scheme plans through switching. Medical scheme plans also attracted low risks to enter and to remain in their risk pools. Attraction of low risks to enter medical schemes was demonstrated with a higher likelihood of households with low risk profiles to join and to switch within schemes to low-risk plan options. South Africa's unregulated medical schemes did not show expected signs of adverse selection.

The technical connection between the expectations for the risk profile variables and the assumed evidence for adverse selection and risk selection was based on previously discussed theory and empirical studies, as already discussed in chapter four. As an example, the higher leaving probability of observed high-risk households, which was interpreted as dumping, is here discussed. It is unlikely that households that display a high-risk profile, i.e. having high claims and essential benefits, have a higher leaving preference. But since the medical schemes can risk select it would be harder for these households to obtain new affordable coverage. Thus it is very likely that the result can be interpreted as evidence for unfavourable risk selection in the form of dumping high risks.

However, here it is rather of interest to discuss whether the results are intuitive and why there is such limited evidence for adverse selection. Using the same example it seems that the results and their interpretation are generally intuitive. First, since risk selection was legal for all medical scheme plans they had a strong incentive to risk select and to stay profitable in this competitive environment. At the same time this executed risk selection strategy prevented any adverse selection in the form of high risks entering plans disproportionately. However, one very interesting fact is that adverse selection effects were observed when they were actually initiated by insurers' risk selection efforts, i.e. low risks left plans of higher-than-average risk or medical scheme plans altogether. Thus the typical adverse selection cycle was not observed. Medical scheme plans experienced a disproportionate number of members with high risk, not because people with the

information advantage of being of higher-than-average risk sought medical scheme plan coverage, but because low risks were enticed to abandon their medical scheme plan coverage through the risk selection efforts of competitors. Also, since there was no evidence for adverse selection, the overlapping problem of adverse selection and moral hazard, widely considered in the literature, does not need to be discussed here.

The main conclusion from this part of the discussion is that the unregulated medical scheme plans encountered risk selection processes but not adverse selection. The absence of adverse selection is one of the most interesting results of this thesis, especially since it is usually the prime focus of the theoretical literature as well as empirical studies. Two possible explanations are discussed here.

First, the studied time period of the unregulated South African medical scheme plans from 1995-1998 coincided with the actual implementation period of new laws that allowed risk selection. Insurance files at the Office of the Registrar showed that few medical scheme plans implemented risk selection earlier than 1993-1994 and most of them only changed their practices at that point in time. The overwhelming evidence for risk selection might then be related to an immediate short-term effect of the deregulation policy on medical scheme plan behaviour. Behaviour adjustment to these changes might take longer on the demand-side (and could be not observed because this case study's short time horizon). This could explain the lack of evidence for adverse selection. Nevertheless, over time, as the option for high risks to obtain affordable medical scheme coverage erodes, they might become more pressed to identify plans into which they can adversely select themselves. However, it is questionable if this would be possible in this setting with the high level of risk selection. It would be interesting to replicate studies of this kind in other unregulated market settings, preferably with more time observations, and assess their evidence for adverse selection.

Second, if the results found here could be replicated for other markets it might be that the problem of asymmetry of information is generally exaggerated and insurers can actually discriminate between risks very well with a few very basic risk indicators. This would be an important finding, because studies in the risk-adjustment literature derived only very limited predictive ability for most of these factors. However, this finding would support the well-know trade-off between efficiency and equity in unregulated health insurance markets. Risk selection here most likely prevents adverse selection and improved market efficiency, but creates inequity issues along the way.

B. Generalisability of results as evidence for increasing inefficiency & inequity

Since a causal relationship between the results of the South African case study and their interpretation as overwhelming evidence for risk selection seems likely, the results should be assessed in light of this thesis' research question. The question here is whether or not it is possible to conclude inequity from the observed results of risk selection.

It seems very likely that inequity occurred, since high risks are more likely to lose affordable health insurance coverage. As pointed out in chapter two, when reviewing the effects of risk selection on market equity there are two equity concerns connected to this result. First, there will be uninsured high risks with low incomes who might be willing to buy insurance, but who cannot afford the offered risk-adjusted medical scheme plan coverage. These high risks might be also excluded as uninsurable risks. Unable to pay for health care out of their pockets they will experience income-related inequity in access to health care. This will have to be combated by South Africa's public sector provisions for the poor and sick, since this option is available. However, South Africa's public health sector provision is of much lower quality than the private health sector. Private sector health care can only be accessed with private or medical scheme plan coverage. Second, there will be uninsured high risks who could afford medical scheme plan coverage but who would be excluded solely on the grounds of being uninsurable high risks. However, in the event of cost intensive ill-health they may be unable to pick up their bills and become a burden for public health sector. Households who should be able to take out insurance coverage privately actually cannot, and have to rely on resources that were allocated to the poor and needy. In conclusion in South Africa's unregulated health insurance market inequity increased due to risk selection.[161]

The research question that motivated this thesis - whether or not adverse selection and risk selection create inefficiency and inequity in unregulated health insurance markets - could be partly answered. Risk selection in unregulated health insurance markets goes hand in hand with inequity. The thesis' results do not only confirm theoretical predictions and intuitive expectations, they also fit into the picture of other research on South Africa's unregulated medical scheme market that used different data and different methods (Hansl and Soderlund 2000, Soderlund and Hansl 1999, 2000). From this discussion it can be also

[161] As pointed out in chapter five and the discussed limitations, the results are hard to interpret in regard to the existence of a separating equilibrium and thus efficiency. It is unclear if risk selection led to an efficient separating equilibrium. However, even if efficiency could be obtained – here in form of a separating equilibrium – charging high risks substantially more is widely seen as inequitable. Anecdotal evidence suggest at least some transitional inefficiency, some plans delaying risk selection strategies experienced the adverse selection 'death spiral' and ended up in bankruptcy.

concluded that the results have policy relevance beyond the South African case study. This thesis' results are generalisable to similar unregulated health insurance arrangements in other country-settings, particularly of low- and middle-income.

However, the generalisability of the thesis's results has some limits. A particular feature of the South African health insurance market was that for-profit insurers operated side by side with the former mutual medical schemes. Their strong competitive position contributed significantly to the observed results of overwhelming risk selection. The existence of such a group of strong competitors might not occur in other cases. However, it can be expected that unregulated health insurance markets are very contested, leading in fact to similar results as those from the South African case study. This predicted observation of risk selection in unregulated health insurance markets might also occur to a different extent in other settings, depending on other factors. Cultural factors (like strong community orientation or high societal cohesion) and economic factors (like income heterogeneity) might mitigate some of the observed effects. For other societies it is in fact expected that the most considerable influence lies in the history and institutional organisation of their health care financing and health care provision. Nevertheless, it is expected that countries will experience predominantly similar effects of risk selection: (1) more competitive/contested health insurance markets and (2) the rise of strong equity concerns.

The general policy relevance of the thesis is based on specific analysis results. It confines the relevancy to the two broad expected effects of risk selection in unregulated health insurance markets, as listed in the previous paragraph. However, general policy strategies to combat these negative effects on equity of less regulation have to be assessed within the context of each specific country in question. In this thesis, specific policy recommendations could only be derived for the South African case study in the context of the unregulated medical scheme market during the observation period of 1994-1998. These policy recommendations will be discussed in the next section.

6.4.2 Policy relevance of the South African case study

The results suggested that the observed South African unregulated medical scheme plans experienced inequity created by risk selection processes.[162] This inequity problem seems to

[162] The transitory inefficiency problem concerned only a few medical plans. Their potential failure could be justified with a higher level of competition after this transition. Thus it would not call for policy interventions.

be of substantial political concern, because high risks cannot obtain or afford health insurance coverage and will stretch the thin resources of South Africa's public health sector that are intended for the poor. It was pointed out that while there is income-related inequity in access to health care there is also inequity in the form of high risks being unable to obtain any insurance coverage, no matter their income level, because they are uninsurable risks. In both of these cases, the equity problem could be overcome if risks would be pooled in larger risk pools where cross-subsidisation makes insurance affordable for all risk types. This could be achieved within a mandatory or social insurance system, for example.

However, there are two specific strategies for political action designed to mitigate the observed deregulation effects in South Africa's medical scheme market that have been proposed and/or implemented. The first is the re-regulation of the medical schemes and second, taking other inequities in South Africa's health sector into account, it is the introduction of a national or social insurance system. These two strategies are discussed below.

First, the next-best solution for the South African regulator could be to simply re-regulate medical schemes and reinstall the rule of the Medical Schemes Act, which prohibited the risk-adjustment of premiums. This would make it easier for high risks to obtain insurance coverage. However, it seems unlikely that this solution would lead to sustainable risk pooling for medical schemes. This is because low risk will still be exposed to the risk selection efforts of the for-profit insurers, who do not fall under the Medical Schemes Act. Thus the equity problem would be solved in the short run, but this strategy might not prevent the financial ruin of many medical schemes in the long run, due to unfair competition from the for-profit insurers. In the long run this would also leave their high-risk ex-clientele later stranded without insurance coverage as potential burdens for the public sector.

A re-regulation of the South African medical scheme market did happen. A new Medical Scheme Act (No. 113) passed the parliament in 1998. The implementation took effect in January 2000, and since then several amendments followed (HST 2003). The new Act opened medical schemes to public enrolment - the enrolment of any eligible applicant is now compulsory. Also, community-rating was reinstated by abolishing risk-adjustments for age and ill-health.[163] The primary objective of the new Act was the

[163] Medical schemes also have to provide a comprehensive package of hospital and outpatient services, the Prescribed Minimum Benefits (PMBs).

extension of coverage in terms of member numbers and their benefits (HST 2003). The opening of the schemes to the large low-income market and the shift of costs to the young and healthy was intended in order to create larger risk pools and bring the costs down for the bulk of the members.

However, the latest report on medical schemes in the *South African Health Review 2002* concludes that these intended consequences did not occur. First, the number of beneficiaries did not significantly increase[164] and medical schemes did not become more affordable for low-income people (HST 2003). Second, for a number of different reasons there was little progress in reducing costs, leading to a persisting trend of increasing contributions. In turn, the risk of low-income earners dropping out of the medical scheme market increased even further, despite the fact that medical scheme contributions were already tax deductible (HST 2003). Furthermore, risk selection practices were not completely eliminated (HST 2003). Thus even after the re-regulation, inequity in terms of health care access remained and the public burden for taking care of the sick and poor increased.

In this respect it also seems very useful to consider the pre-existing inequities in South Africa's health care financing and provision. The medical scheme market only covers the 20% of the South African population, which is able to afford health insurance. An even smaller proportion, 4% of the population that is insured with for-profit insurers, while the larger population relies on a tax-base financed public system (HST 2003). In order to tackle the persisting inequities after the end of the Apartheid state, social health insurance became a widely discussed alternative in South Africa's public, academic, and political arena (McIntyre and Owen 1994, DoH 1997a, HST 2003). Mandatory social insurance could create larger risk pools with better cross-subsidisation and solve the pressing equity issues (n/e/r/a 1996).

A Social Health Insurance (SHI) plan was indeed put forward by the Department of Health (DoH 1997). The discussion of alternative forms of health care financing continues and, in particular, SHI is still a priority among interventions for addressing equity, efficiency and long run sustainability problems in South Africa's health care financing (HST 2003).

Both of the two political strategies described could have been derived as policy recommendations from the thesis' results. This only confirms its policy relevance in answer to the research question. However, the re-regulation of the medical schemes,

effective by 2000, has been, to date, not successful. This is the case because this intervention only considers the medical scheme plans and does not prevent for-profit insurers from practicing risk selection. This biased selection on one end of the South African health insurance markets breaks, and will continue to break, the cross-subsidising risk pooling arrangements of the medical scheme market. In order to prevent inequity and to protect the public sector from becoming more and more responsible for taking care of the sick and poor, it is recommended here that the regulatory framework be extended beyond the medical scheme market to all private health insurance arrangements in South Africa. This would protect the medical scheme market from loosing low risks to the for-profit market and enable more sustainable risk pooling. However, this would not solve the bigger inequity issue related to South Africans' unequal access to health care services (HST 2003). Here the social health insurance plan, put forward by the Department of Health in 1997 as an alternative way of financing health care, seems a more comprehensive approach. It not only has the potential to solve the equity problems pointed out in this thesis, since larger risk pools provide better cross-subsidisation, but it also has the potential to provide a sustainable long-term solution for South Africa's health care financing.

The next section will summarise this discussion chapter.

6.5 Summary

This chapter assessed the thesis' research question, whether adverse selection and risk selection create inefficiency or inequity in unregulated health insurance markets, by discussing the thesis' main contributions, selected limitations and policy relevance.
Finally, with the results of this thesis' empirical analysis it was possible to discuss this thesis' policy relevance and answer parts of the research question, concluding that risk selection creates inequity.

The second part of this chapter discussed application of the group method to an empirical investigation in a middle-income country as the thesis' main contribution. The performed empirical analysis contributes to the few empirical studies that investigate selection processes in low- and middle-income countries.

[164] Between 2000 and 2001 the membership increased by less than one percent (HST 2003).

Part three discussed a number of limitations, which were mainly identified while conducting the empirical analysis into South Africa's unregulated medical scheme market. In addition to technical limitations, the two fundamental limitations of research of this kind, namely the assumption of rationality and the inference of motivation from observed outcomes of behaviour were discussed. The technical limitations were related to the case study data and the empirical analysis performed in this thesis.

The fourth part of this chapter discussed the thesis' policy relevance. First, the general political relevance of this thesis' results was confirmed, since causality between the analysis results and their interpretation as evidence for risk selection is very likely. The results partly answered the research question that motivated this thesis: risk selection creates in unregulated health insurance markets inequity. This result seem generalisable to similar settings in other countries. Second, the specific policy relevance of the thesis' empirical results for the South African case study was discussed. Two proposed and/or implemented strategies for political action in order to mitigate the rising equity concern for South Africa's medical scheme market were presented.

Chapter 7

Summary & conclusion

7.1 Principal findings

The purpose of this section is to summarise and conclude the principal findings of the thesis. The objective of the thesis was to conduct an empirical analysis that would determine whether unregulated health insurance markets experience adverse selection and risk selection. From this, two research aims were derived. The first research aim was to derive an appropriate method for empirical investigations into adverse selection and risk selection that would use available quantitative techniques and easily accessible data. The second research aim was to apply this method to an empirical analysis of case study data from an unregulated health insurance market in a middle-income country. This empirical analysis was the heart of the thesis, which was motivated by the research question: are unregulated health insurance markets characterised by adverse selection and/or risk selection, thereby creating inefficiency or inequity? This section will first summarise the thesis in regard to its research objective and research aims. Second, the answer to the research question will be summarised.

7.1.1 Research objective and research aims

The purpose of this section is to summarise the usefulness of the group method and to summarise the results of the South African case study to which the group method was applied.

A. The group method for empirical investigations

The group method is based on the argument that the explanation or illustration of selection processes using the group dimension is equally fitting and revealing as

compared to the previously emphasised time dimension. Previous empirical investigations examined characteristics of individuals, who are the carriers of the selection processes, by dividing data of insured households into subgroups of households who move into, out of, and between insurance plans. They observed whether households with certain risk profiles are more prominent in certain mover groups and in certain risk pooling plans. From this group method it was possible to derive an empirically testable hypothesis, this thesis' main research hypothesis: South Africa's unregulated medical schemes encountered a lack of market-wide risk pooling and a lack of risk pooling within the market (between insurers' plan options) due to adverse selection and risk selection.

Matching quantitative techniques were identified that are able to explain group memberships for households using a multivariate risk profile, i.e. test whether or not households of different mover groups have significantly different risk profiles as compared to a control group. The risk composition of groups can be then attributed to certain selection processes. Discriminant analysis and, in particular, logistic regression analysis were identified as applicable quantitative techniques for empirical investigations because of their ability to evaluate group differences on a multivariate profile.

The group method is applicable to cross-sectional or panel data of insured individuals or households that contain the dates when these people join, leave, and switch insurance plans as well as basic membership and insurance utilisation information. This data is widely available because insurance companies routinely collect it. This data also represents the main limitation of the group approach, because real world data is often flawed or even non-existent. However, if the data limitation problem can be overcome, the group method is generalisable to different country-settings, various insurance arrangements and also non-health related markets.

It can be concluded here that the first research aim was fulfilled - the derived group method is highly appropriate for empirical investigations into adverse selection and risk selection. The validity of the group method for empirical investigations was tested in this thesis on real world data of South Africa's unregulated medical schemes.

B. Result summary for the South African case study

The empirical analysis performed in the thesis, applying the group method, led to model estimation results from which the existence of selection processes in the South African health insurance market during the time of observation was concluded. The analysis results and their interpretation, based on three research hypotheses that were derived

from this thesis' main hypothesis, led to three conclusions. First, overwhelming evidence for favourable risk selection of low risks into medical scheme plans could be derived. Second, the existence of evidence for unfavourable risk selection by medical scheme plans in the form of dumping high risks was concluded. Third, leavers have some low risk characteristics, which seems like evidence for adverse selection. However, since the typical adverse selection cycle was not observed, it was concluded that this result actually detected external favourable risk selection by the for-profit medical scheme plan competitors.

In conclusion of the assessment of these research subhypotheses, the main research hypothesis, that South Africa's unregulated medical scheme market encountered a lack of risk pooling, could be verified. The thesis' main result was that risk selection, not adverse selection, contributed to a lack of risk pooling. First, the overall medical scheme plan lacked risk pooling due to external favourable risk selection by for-profit medical scheme plan competitors. Competition between the medical scheme plans and for-profit insurers caused risk selection, breaking established risk pooling arrangements. Second, there was a lack of risk pooling between different medical scheme plans. Competition here caused risk selection between medical scheme plans. Over the four-year observation period high-risk plan options continued to pool high risks and low-risk plan options continued to pool low risks.

It was concluded that the second research aim was fulfilled - the application of the group method to an empirical investigation into selection processes using data from South Africa's unregulated medical scheme market produced distinguishable evidence for selection processes.

Both research aims - the development of an easily applicable method for empirical investigations and testing of this method on a case study data set - were effectively accomplished.

7.1.2 Research question and policy relevance

It is the purpose of this section to present the conclusion for the thesis research question and confirm the limits of this thesis' scope, particularly regarding policy relevance. The results from the empirical analysis partly supported the research question that motivated this thesis. While it could be confirmed that risk selection created inequity, it could be not

concluded that adverse selection created inefficiency (because no adverse selection evidence was found).

The general political relevance of the thesis' results was concluded from the very likely causality between the case study's analytical results and expected effects of risk selection. The result of inequity due to risk selection was deemed generalisable to similar settings in other countries. It was predicted that unregulated health insurance markets in other countries would most likely experience two of the observed risk selection effects: (1) more competitive/contested health insurance markets and (2) the rise of strong equity concerns, varying in their extent, depending on economic, cultural, or institutional factors.

However, the policy relevance is limited to these expected *effects* of risk selection, and the result that less regulation in health insurance markets cannot resolve or address all issues that health care systems in middle- and low-income countries face. This is particularly true of inequity problems (resulting from risk selection or not). General strategies to combat these negative equity effects of less regulation have to be assessed within the context of each specific country. Specific policy recommendations were derived in this thesis, for the South African case study. The case study context placed the results of the analysed data into the time period 1994-1998, when South Africa's medical scheme plans were less regulated and could risk-relate premium contributions.

Two specific policy implications for the South African case study that consider the broader features of the South African health sector were brought forward. Both policies essentially aim towards an enlargement of health insurance risk pools in order to make insurance more affordable for low-income people and in order to lighten the public burden for taking care of the sick and poor. The first is to impose more regulation, not less, on the medical scheme market and extend the regulatory framework beyond the medical scheme market to all private health insurance arrangements in South Africa. Part of this recommendation was implemented with the re-regulation of South Africa's medical scheme market, effective from 2000. However, this re-regulation did not consider the practice of for-profit insurers. As a result it could be observed that biased risk selection by South Africa's for-profit insurers continued to break the cross-subsidising risk pooling arrangements of the medical scheme plans by extracting their low risks. The second recommendation concerns the proposed implementation of some kind of mandatory health insurance system, such as the discussed Social Health Insurance plan. The later recommendation addresses the larger equity issues in South Africa's health care system. It

would not only create larger risk pools with better cross-subsidisation, it would also provide a sustainable long-term solution for South Africa's health care financing.

7.2 Limitations

This thesis' economic approach accepted the assumption of the general model of rational economic behaviour as its analytical context and the empirical analysis drew inference of motivation from observed outcomes of behaviour. These are also the two fundamental limitation of research of this kind, besides the technical limitations of the actual data and analysis, as discussed in chapter six. However, it was concluded that rationality can be precise and prescriptive, allowing the economic approach to complement political science, sociology, psychology and other disciplines. The issue of making inferences about behavioural motivations based on observed consequences or outcomes of behaviour is a limitation that empirical research in all disciplines faces. Scientific approaches are built on interplay between theory and observable data. However, technical data and analysis limitations can influence these inferences.

As a retrospective study, data was limited to what had been recorded and to data that were not designed for assessing risk pooling and selection processes. However, it was concluded that the two main data limitations - first, the practical data subgroup & plan classification and second, the limits of the variables used as risk factors - were sufficiently considered to ensure that the statistical analysis results were still valuable. Also, the data set contained a given set of medical scheme plans, and although it was assumed that this sample was representative, this may not have been the case, which could have biased the results. Another weakness was the lack of control for environmental or organisational factors, such as plan management details that might have affected plan risk pooling.

Technical limitations related to the analysis, such as limits on the model results interpretations and the model fit assessment, were acknowledged and, if possible, taken into consideration. The lack of adequate model fit assessment methods for the thesis' correlated data models seemed the most serious limitation here. This made it hard to evaluate the quality of the models in terms of efficiency and predictive ability. However, a good model is not made of only descriptive capacity and predictive ability, although models are preferred that can be successfully taken to new data and that can help answer

policy questions. Arguably, these data and analysis limitations influenced the offered results and their interpretation as behavioural outcomes from specific motivations.

One important limitation was that the result of the analysis could be not very well interpreted within the framework of a separating equilibrium. A longer time-dimension of the data and the analysis of time-trends might have lent itself better to the assessment of a separating equilibrium hypothesis.

7.3 Contribution to knowledge

In spite of the limitations discussed in chapter six the thesis contributes to the empirical literature on adverse selection and risk selection in health insurance markets. Adverse selection and risk selection are of great concern both in academic and political circles. Most literature on these issues is focused on high-income countries. However, there are a few studies in the context of low- or middle-income countries (e. g. Sapelli and Vial 2003, Soderlund and Hansl 2000). The thesis' empirical analysis of selection processes in South Africa's medical scheme market adds to these few existing studies, thereby increasing the understanding of the behaviour of health insurance market participants in a middle-income country context.

Empirical evidence for adverse selection and risk selection is hard to identify. The empirical analysis was the centrepiece of this thesis and its major contribution. The significance of this thesis lies in the successful application of the group method to the empirical analysis of the South African case study.

The empirical analysis of the South African case study data adds to the few empirical studies that explore adverse selection and risk selection in low- and middle-income countries. It contributed in particular to the few studies that have tried to empirically identify adverse selection and risk selection in South Africa's medical scheme market (Soderlund and Hansl 1999, 2000, Hansl and Soderlund 1999). However, none was based on individual-level insurance data as used in this thesis, or applied the specific group method.

In addition to the empirical analysis, this study also explored the policy relevance of the results. Specifically, the impact of less regulation for the South African medical scheme market was examined. The results indicated that the effects of less regulation in middle-income country health sector reforms, in particular, leaving health insurance arrangements unregulated, conflicts with the common health policy objective of equity.

More competition and efficiency comes at the price of less equity in health care access for low-income groups, confirming the known efficiency-equity trade-off. The thesis' policy recommendations support the position that there is need for more, not less, regulation in low- and middle-income countries in order to make health care financing and health care provision more equitable.

7.4 Outlook for future research

This thesis explored the research question of whether unregulated health insurance markets are characterised by adverse selection and/or risk selection, thereby creating inefficiency or inequity. The focus was on an empirical investigation into evidence for these selection processes. However, this thesis concludes with a call for more empirical studies of adverse selection and risk selection and their effects. First, studies of this kind are missing, in particular, for the health sector in low- and middle-income countries. Second, the thesis' result, which showed a lack of evidence for adverse selection, needs to be replicated and confirmed with other studies. The result seems controversial since it suggests that the problem of asymmetry of information in health insurance markets might be exaggerated in theoretical and empirical literature. It seems that insurers can actually risk discriminate (and risk select) very well with only a few basic risk indicators. However, the limited time horizon of this thesis also limited its ability to test explicitly for a separating equilibrium. It would be interesting to replicate studies of this kind in other unregulated market settings, preferably with more time observations, and assess their evidence for adverse selection and a separating equilibrium. Aside from this, there are many related research questions that were not answered here.

For example, while there is a large body of regulation theory (which was not the subject of this thesis) there are remarkably few researchers who explore the particular issue of health sector regulation in low- and middle-income countries (see Kumaranayake et al. 2000 and Hongore and Kumaranayake 2000 for good examples to the contrary). This seems particularly astonishing, since these countries can ill afford to waste resources that will not achieve all of the proclaimed aims of health sector regulation. This concerns potential future research questions, which assess the effects of more or less regulation in the health care sector. Equally important research questions would be ones which would compliment these effect analyses with decision analyses concerning the political decision

making process about whether to increase or decrease regulation in low- or middle-income countries' health sectors. There are several explanations that suggest themselves as to why there is such limited literature on these issues, and from these, several areas for future research and political engagement can be easily formulated.

First, there is the need for more and better data in order to conduct assessments of the effects of implemented policies. It would be helpful to not only provide low- and middle- income countries with policy recommendations but also with specific monitoring tools for assessing the effects of each recommended policy. In particular, tools should be developed that allow easy data collection for before-and-after scenario comparisons. There is a lack of easily applicable benchmarks that evaluate policies in those countries.

Second, alternative health care financing methods have to be further explored for low- and middle- income countries that have limited or no history of Western health insurance arrangements. This is a particular political imperative since many public health care sectors in low- and middle- income countries are at the limit of their capacities, if existing at all. Here sustainable long-term solutions, rather than political patchwork, as in the South African re-regulation example, are called for. Moreover, these alternatives have to take the limits of political institutions in those countries into account. This includes, for instance, their often moderate ability to perform complicated regulation tasks that Western governments also struggle with (see, for example, on government failure: Le Grand 1991). Until now, research and experiments on this issue have been sparse for low- and middle- income countries (recent exceptions are: Wiesmann and Juetting 2000, Purohit 2001, McCord 2001, Barninghausen and Sauerborn 2002 and Miriam 2003).

Finally, aside from conveniently demanding greater commitment from politicians in low- and middle- income countries, who are often trapped and challenged by electoral cycles and bureaucratic institutions comparable to the West, economists are asked to examine their approaches to understanding and explaining behaviour of participants in health insurance markets. For example, Nyman's *Theory of Demand for Health Insurance* (2003) comprises a new approach that is truly interested in understanding behaviour related to health insurance coverage, rather than focusing purely on mathematical aspects of modelling that behaviour.[165] He develops a health insurance demand theory based on the fact that consumers are not necessarily averse to risk or uncertainty but are averse to

[165] He argues that understanding health insurance demand does not require complex models but experientially derived intuition (Nyman 2003).

income losses.[166] This approach can explain the intuitive conclusion that the voluntary purchase of fair health coverage makes consumers better off and not worse as controversially discussed among economists (e.g. Pauly 1968, Pauly 1983). Thus economists who are interested in empirical modelling of health insurance behaviour should not apply the conventional theoretical model without scrutiny, simply because it allows them to perform the desired empirical analysis. At the least, empirical studies that try to model insurance behaviour should explicitly state or clarify behavioural assumptions that underlay their modelling efforts.

[166] Thus insurance is demanded because of its income implications, not because of implications regarding certainty (Nyman 2003).

Bibliography

Abbring, J.H., Chiaporri, P.-A., Heckman, J.J. and Pinquet, J. (2003a): Adverse selection and moral hazard in insurance: Can dynamic data help to distinguish? *Journal of the European Economic Association* 1:2: 512-521.

Abbring, J.H., Chiaporri, P.-A., Heckman, J.J. and Pinquet, J. (2003b): Moral hazard and dynamic insurance data. *Journal of the European Economic Association* 1:4: 746-820.

Akerlof, G. (1970): The market for lemons: Qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics* 84:488-500.

Allard, M., Creste, J.-P. and Rochet, J.-C. (1998): Pooling and separating equilibria in insurance markets with adverse selection and distribution costs. *Mathematics and Economics* 22:3:300-300.

Altman, D., Cutler, D.M. and Zeckhauser, R.J. (1998): Adverse selection and adverse retention. *American Economic Review, Papers and Proceedings* 88:2:122-126.

Arnott, R., Greenwald, B. and Stiglitz, J.E. (1994): Information and economic efficiency. *Information Economics & Policy* 6:1:77-88.

Arrow, K.J. (1963): Uncertainty and the welfare economics of medical care. *American Economic Review* 58:5:941-973.

Ash, A., Porell, F., Gruenberg, L., Sawitz, E. and Beiser, A. (1989): Adjusting Medicare capitation payments using prior hospitalization data. *Health Care Financing Review* 10:4:17-29

Barnett, V. and Lewis, T. (1984), *Outliers in Statistical Data*, 2nd ed. New York.

Barnighausen, T. and Sauerborn, R. (2002): One hundred and eighteen years of the German health insurance system: are there any lessons for middle- and low-income countries? *Social Science and Medicine* 54:10:15559-1587.

Barr, N. (1998): *The economics of the welfare state.* 3rd edition. Oxford.

Barros, P.P. (2003): Cream-skimming, incentives for efficiency and payment systems. *Journal of Health Economics* 22:3:419-443.

Bartholomew, D.J. (1987): *Latent variable models and factor analysis.* Griffin Statistical Monographs and Courses Vol.40, London.

Beck, K. and Zweifel, P. (1998): Cream-skimming in deregulated social health insurance: evidence from Switzerland. In: Zweifel, P. (ed.): *Health, the medical profession, and regulation.* Dordrecht, pp.211-227.

Becker, G.S. (1976): *The economic approach to human behavior.* Chicago.

Belli, P. (2001): How adverse selection affects the health insurance market. *Policy research working paper 2574.* World Bank, Washington.

Berliner, B. (1982): *Limits of insurability of risk.* Upper Saddle River.

Berliner, B. (1984): Some thoughts on uncertainty, the transfer of knowledge and insurability. *The Geneva Papers on Risk and Insurance* 9:33:380-395.

Berman, P., Sisler, D. and Habicht, J. (1989): Equity in the public sector primary health care: the role of service organization in Indonesia. *Economic Development and Cultural Change* 37:4:777-803.

Besley, T., Hall, J. and Preston, I. (1999): The demand for private health insurance: do waiting lists matter? *Journal of Public Economics* 72:155-181.

Betranou, F. (1998): Health care services utilization and health insurance coverage: evidence from Argentina. *Revista de Análisis Económico* 13:2:25-52.

Bice, T.W. (1975): Risk vulnerability and enrollment in prepaid group practice. *Medical Care* 13:8:698-703.

Billi, J.E., Wise, C.G., Sher, S.I., Duran-Arenas, L., and Shapiro, L. (1993): Selection in a preferred provider organization enrolment. *Health Services Research* 28:5:563-75.

Blankart, C.B. (1998): *Oeffentliche Finanzen in der Demokratie. Eine Einfuehrung in die Finanzwissenschaft.* Muenchen.

Blumberrg, L.J. and Nichlos, L.M. (1996): First, do no harm: Developing health insurance market reform packages. *Health Affairs* 15:3:35-53.

Bozeman, B. and Bretschneider, S. (1986): Public management information systems: Theory and prescription. *Public Administration Review* 46:Special Iss (Nov):478.

Brown, R.S. et al. (1993): Do HMOs work for Medicare? *Health Care Financing Review* 15:1:7-23.

Brown, H. and Prescott, R. (1999): *Applied mixed models in medicine.* Statistics in Practice Series, England.

Browne, M.J. (1992): Evidence of adverse selection in the individual health insurance market. *Journal of Risk & Insurance* 59:1:13-33.

Browne, M.J. and Doerpinghaus, H. (1993): Information asymmetries and adverse selection in the market for individual medical expense insurance. *Journal of Risk & Insurance* 60:2:300-312.

Browne, M.J. and Doerpinghaus, H. (1994): Asymmetric information and the demand for Medigap insurance. *Inquiry* 31:4:445-450.

Buchmueller, T.C. and Feldstein, P.J. (1997): The effect of price on switching among health plans. *Journal of Health Economics* 16:321-247.

Burchardt, T. and Propper, C. (1999): Does the UK have a private welfare class? *Journal of Social Policy.* 28:4:643-665.

Cameron, A.C., Trivedi, P.K., Milne, F. and Piggot, J. (1988): A microeconomic model of the demand for health care and health insurance in Australia. *Review of Economic Studies* 55:85-106.

Cardon, J. and Hendel, I. (1996): Asymmetry of information in health care and health insurance markets: evidence from the National Medical Expenditure Survey", mimeo.

Cartwright, W.S.; Hu, T.-W., and Huang, L.-F. (1992): Impact of varying Medigap insurance coverage on the use of medical services for the elderly. *Applied Economics* 24:5:529-539.

Chernichovsky, D. and van de Ven W. (2003): Risk adjustment in Europe. *Health Policy* 65:1:1-3.

Chiaporri, P.-A. and Salanie, B. (2003): Testing contract theory: A survey of some recent work. In: Dewatripont, M., Hansen, L. and Turnovsky, P. (eds.): *Advances in Economics and Econometrics - Theory and Applications, Eighth World Conference,* Econometric Society Monographs. Cambridge pp.115-149.

Chinitz, D. and Cohen, J. (eds.) (1998): *Governments and health systems--Implications of differing involvements.* New York.

Clark, J.B. (1893): Insurance and business profit. *Quarterly Journal of Economics* 7:40-54.

Conrad, D.A., Grembowski, D. and Milgrom, P. (1985): Adverse selection within dental insurance markets. In: Scheffler, R.M. and Rossiter, L.F. (eds.): *Advances in health economics and health services research* Vol 6. Greenwich, pp.171-190.

Constitution of the Republic of South Africa, Act 200 1993. Available online: http://www.polity.org.za/html/govdocs/legislation/1993/constit0.html

Cremer, J. and Kahlil, F. (1992): Gathering information before signing a contract. *American Economic Review* 82:3:566-578.

Culyer, A.J., and Newhouse, J.P. (eds.) (2000): *Handbook of Health Economics.* Amsterdam.

Cutler, D. (1994): *Market failure in small group health insurance.* NBER Working Paper No. 4879. Palo Alto.

Cutler, D. and Reber, S. (1998): Paying for health insurance: the trade-off between competition and adverse selection. *The Quarterly Journal of Economics* 113:2:433–466.

Cutler, D. and Zeckhauser, R. (1998): Adverse selection in health insurance. In: Garber, A. (ed.): *Frontiers in health policy research Vol. 1.* NBER, Palo Alto.

Cutler, D. and Zeckhauser, R. (2000): The anatomy of health insurance. In: Cuyler, A. J. and Newhouse, J.P. (eds.): *Handbook of health economics. Vol. 1A.* Amsterdam pp. 563-643.

Dahlby, B. (1981): Adverse selection and pareto improvements through compulsory insurance. *Public choice* 37:548-558.

Daniels, N. (1982): Equity in access to health care: some conceptual and ethical issues. *Milbank Quarterly* 60:1:51-81.

Danzon, P.M. (2002): Welfare effects of supplementary insurance: a comment. *Journal of Health Economics* 21:5:923-926.

Davis, C.S. (2002): *Statistical methods for the analysis of repeated measurements.* New York.

Dempster, A.P. and Rubin, D.B. (1983): Overview In: Madow, W.G., Olkin, I. and Rubin, D.B. (eds.): *Incomplete data in sample surveys: Theory and bibliography.* New York.

Department of Health (1996): *Restructuring the national health system for universal primary health care.* Main Report 2. Report of the Committee of Inquiry into National Health Insurance System. Pretoria.

Department of Health (1997a): *A social health insurance scheme for South Africa.* Policy document. Pretoria.

Department of Health (1997b): *The White paper on the transformation of health care system in South Africa.* Pretoria.

Department of Health (2001): *Annual Report 2000-2001.* Pretoria.

Dewatripont, M., Hansen, L. and Turnovsky, P. (eds.) (2003): *Advances in Economics and Econometrics - Theory and Applications, Eighth World Conference,* Econometric Society Monographs. Cambridge.

Diehr, P., Madden, C.W., Martin, D.P., Patrick, D.L., Mayers, M.; Char, P., Skilman, S., Cheadle, A., Fishman, P., Hoare, G. et al. (1993): Who enrolled on a state program for the uninsured: was there adverse selection? *Medical Care* 31:12:1093-105.

Diggle, P.J. and Kenward, M.G. (1994): Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* 43:49-93.

Diggle, P.J., Heagerty, P.J., Liang, K.Y. and Zeger, S.L. (2002): *Analysis of longitudinal data.* Oxford.

Dowd, B. and Feldman, R. (1985): Biased selection in twin cities health plans. In: Scheffler, R.M. and Rossiter, L.F. (eds.) *Advances in health economics and health services research* Vol 6. Greenwich. pp.127-147.

Dowd, B., Feldman, R., Moscovice, I., Wisner, C., Bland, P. and Finch, M. (1995): An Analysis of selectivity bias in the Medicare AAPCC. *Health Care Financing Review* 17:3:35-59.

Eggers, P.W. and Prihoda, R. (1982): Pre-enrollment reimbursement patterns of Medicare Beneficiaries enrolled in 'at-risk' HMOs. *Health Care Financing Review* 4:1:55-73.

Eggers, P.W. (1980): Risk differential between Medicare beneficiaries enolled and not enrolled in an HMO. *Health Care Financing Review* Winter:91-99.

Ellis, R.P. (1985): The effect of prior-year health expenditures on health coverage plan choice. In: Scheffler, R.M. and Rossiter, L.F. (eds.): *Advances in health economics and health services research* Vol 6. Greenwich, pp.127-147.

Ellis, R.P. (1989): Employee choice of health insurance. *Review of Economics and Statistics* 71:2:215-223.

Ellis, R.P., et al. (1995): Diagnosis-based risk adjustments for Medicare capitation payments. *Health Care Financing Review* 17:3:101-129.

Ellis, R.P. (1998): Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics* 17:5:537–556.

Enthoven, A. and Singer, S. (1995): Market-based reform: What to regulate and by whom *Health Affairs* 14:1:105-119.

Epstein, A.M. and Cumella, E.J. (1988): Capitation payment: Using predictors of medical utilisation to adjust rates. *Health Care Financing review* 10:1:51-69.

Farley, P.J. and Monheit, A.C. (1985): Selectivity in the demand for health insurance and health care. In: Scheffler, R.M. and Rossiter, L.F. (eds.): *Advances in health economics and health services research* Vol 6. Greenwich, pp.231-252.

Feldman, R. and Dowd, B. (2000): Risk segmentation: Goal or problem? *Journal of Health Economics 19:499-512.*

Frank, R. G., Glazer, J. and McGuire, T. G. (2000): Measuring adverse selection in managed health care. *Journal of Health Economics 19:6:829-854.*

Frey, B.S. (1980): *Modern political economy.* Oxford.

Garber, A. (ed.) (1998): *Frontiers in health policy research Vol. 1.* NBER, Palo Alto.

Gravelle, H., Sutton, M., Morris, S. et al. (2003): Modelling supply and demand influences on the use of health care: implications for deriving a needs based capitation formula. *Health Economics* 12:985-1004.

Greene, W.H. (2002): *Econometric analysis.* New York.

Griffith, M.J. and Baloff, N. (1981): Membership duration and utilisation in a prepaid group practice. *Medical Care* 19:12:1195-210.

Gruenberg, L., Kaganova, E. and Hornbrook, M.C. (1995): Improving the AAPCC with health-status measures from the MCBS. *Health Care Financing Review* 17:3:59-77.

Government of South Africa (1999): *South Africa Yearbook 1998-1999.* Government Communication and Information Systems (GCIS). Pretoria.

Government of South Africa (2003): *South Africa Yearbook 2002-2003.* Government Communication and Information Systems (GCIS). Pretoria

Gromping, U. (1996): A note on fitting a marginal model to mixed-effects log-linear regression data via GEE. *Biometrics* 52:1:280-285.

Grossman, H.I. (1979): Adverse selection, dissembling, and competitive equilibrium. *Bell Journal of Economics* 10:1:336-343.

Guillen, A.M. (2002): The politics of universalisation: establishing national health services in Southern Europe. *West European Politics* 25:4:49-68.

Hadi, A.S. (1992): Identifying multiple outliers in multivariate data. *Journal of Royal Statistical Societ,* Series B 54:761-771.

Hadi, A.S. (1992): A modification of a method for the detection of outliers in multivariate samples. *Journal of Royal Statistical Society,* Series B 56:393-396.

Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, W.C. (1998): *Multivariate data analysis.* 5th ed. New Jersey.

Hansen, B.O. and Keiding, H. (2002): Alternative health insurance schemes: a welfare comparison. *Journal of Health Economics* 21:5:739-756.

Hansl, B. and Soderlund, N. (1999): *Determinants of private health insurance coverage in a middle-income country - South Africa.* Paper presented at the 2nd World Conference of the International Health Economics Association (iHEA) in Rotterdam.

Hawley, F.B. (1901): Enterprise and profit. *Quarterly Journal of Economics* 15:75-05.

Haynes, J. (1895): Risk as an economic factor. *Quarterly Journal of Economics* 9:409-449.

Health Systems Trust (1996): *South African Health Review 1995.* Durban.

Health Systems Trust (1998): *South African Health Review 1997.* Durban.

Health Systems Trust (2003): *South African Health Review 2002.* Durban.

Hill, J.W. and Brown, R.S. (1990): Biased selection in the TEFRA HMO/CMP program. *Mathematica Policy Research* (MPR Reference No 7786-503).

Hisao, C. (1986): *Analysis of panel data.* Econometric Society Monographs No. 11. Cambridge.

Holahan J., Nichols L.M., Blumberg L.J. and Shen Y.-C. (2003): A new approach to risk-spreading via coverage-expansion subsidies. *The American Economic Review* 92:2:277-282(6).

Hongore, C. and Kumaranayake, L. (2000): Do they work? Regulating for-profit providers in Zimbabwe. *Health Policy and Planning* 15:4:368-377.

Hopkins, S. and Kidd, M.P. (1996): The determinants of the demand for private health insurance under Medicare. *Applied Economics* 28:12:1623-1632.

Hosmer, D.W. and Lemeshow, S. (2000): *Applied logistic regression.* 2nd ed. Wiley Series in Probability and Statistics. New York.

Hougaard, P. (2000): *Analysis of multivariate survival data.* Statistics for Biology and Health Series. New York.

Hurd, M.D. and McGarry, K. (1997). Medical insurance and the use of health care services by the elderly. *Journal of Health Economics* 16:129-154.

International Monetary Fund (1997): *Uganda.* Enhanced Structural Adjustment Facility Policy Framework Paper 1997/98-1999/00. Available online: http://imfnt1x.imf.org/external/np/pfp/uganda/uganda00.htm

International Monetary Fund (1998a): *Cameroon.* Enhanced Structural Adjustment Facility Medium-Term Economic and Financial Policy Framework Paper (1998/99-2000/01). Available online: http://www.imf.org/external/np/pfp/camer/camtoc.htm

International Monetary Fund (1998b): *Burkina Faso.* Enhanced Structural Adjustment Facility Policy Framework Paper for 1998-2000. Available online: http://www.imf.org/external/np/pfp/bf/bfasoind.htm

Jackson-Beeck, M. and Kleinman, J.H. (1983): Evidence for self-selection among health maintenance organisation enrollees. *Journal of the American Medical Association* 250:20:2826-2829.

Juba, D.A., Lave, J.R. and Shaddy, J. (1980): An analysis of the choice of health benefits plans. *Inquiry* 17:62-71.

Kahnemann, D., Slovic, P. and Tversky, A. (eds.) (1982): *Judgement and uncertainty: Heuristic and biases.* Cambridge.

Karplus, M. and Betranou, F.M. (1999): Are market-oriented health insurance reforms possible in Latin America? - The cases of Argentina, Chile and Columbia. *Health Policy* 47:1:19-36.

Keenan, P., Buntin, M.B., McGuire, T. and Newhouse, J., (2001): The prevalence of formal risk adjustment in health plan purchasing. *Inquiry* 38:245-259.

Kifmann, M. (2002): Community rating in health insurance and different benefit packages. *Journal of Health Economics 21:5:719-738.*

King, D.R. and Mossialos, E. (2002): *The determinants of private medical insurance prevalence in England.* LSE Health and Social Care Discussion Paper. London School of Economics and Political Science.

Kirchgaessner, G. (1991): *Homo economicus. Das oekonomische Modell individuellen Verhaltens und seine Anwendungen in den Wirtschaftswissenschaften.* Tuebingen.

Knight, F.H. (1921): *Risk, uncertainty and profit.* Boston: Houghton Mifflin, 1945.

Kronick, R., Dreyfus, T., Lee, L. and Zhou, Z. (1995): Diagnostic risk adjustments for Medicaid: the disability payment system. *Health Care Financing Review* 17:3:7-35.

Kumaranayake, L. (1997): The role of regulation: Influencing private sector activity within health sector reforms. *Journal of International Development* 9:4:641-649.

Kumaranayake, L. (1998): *Incentives and incentive regulation in the health sector: Perspectives from low- and middle-income countries.* Unpublished Paper of the Quasi-Market Research Seminar in April 1998.

Kumaranayake, L., Lake, S., Mujinja, P., Hongore, C. and Mpembini, R. (2000): How do countries regulate the health sector? Evidence from Tanzania and Zimbabwe. *Health Policy and Planning* 15:4:357-367.

Lacker, J.M. (1994): Does adverse selection justify government intervention in loan markets? *Economic Quarterly* 80:Winter:61-95.

Laird N.M. and Ware, J.H. (1982): Random-effects models for longitudinal data. *Biometrics*, 38:963-974.

Lamers, L.M., van Vliet, R.C. (1996): Multiyear diagnostic information from prior hospitalization as a risk-adjuster for capitation payments. *Medical Care* 34:6:549-61.

Langwell, K.M and Hadley, J.P. (1989): Evaluation of the Medicare competition demonstration. *Health Care Financing Review* 11:2:65-80.

Le Grand, J. (1991): The theory of government failure. *British Journal of Political Science* 21:423-442.

Lee, K. (1991): Transaction costs and equilibrium pricing of congested public goods with imperfect information. *Journal of Public Economics* 45:3: 337-362

Liang, K.Y. and Zeger, S.L. (1986): Longitudinal data analysis using generalized linear models. *Biometrika* 73:12-22.

Liang, K.Y. and Zeger, S.L. (1993): Regression analysis using for correlated data. *Annual Review of Public Health*, 14:43-68.

Little, R.J.A. (1995): Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* 90:1112-1121.

Little, R.J.A. and Rubin, D.B. (2002): *Statistical analysis with missing data.* 2nd ed. New York.

Long, S.H., Settle, R.F., and Wrightson, C.W. (1988): Employee premiums, availability of alternative plans, and HMO disenrollment. *Medical Care* 26:10:927-938.

Long, J.S. and Freese, J. (2001): *Regression models for categorical dependent variables using STATA.* STATA Press.

Macho-Stadler, I. and Perez-Castrillo, D. (1997): *An introduction to the economics of information. Incentives & contracts.* Oxford.

Madow, W.G., Olkin, I. and Rubin, D.B. (eds.) (1983): *Incomplete data in sample surveys: Theory and bibliography*. New York.

Manning, W.G., Newhouse, J.P., Duan, N., Keeler, E.B., Leibowitz, A. and Marquis, M.S. (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. *American Economic Review* 77:3:251-277.

Manning, W.G. and Marquis, M.S. (1996): Health insurance: the trade-off between risk-pooling and moral hazard. *Journal of Health Economics* 15:5:609-639.

McGuire, T.G. (1981): Price and membership in a prepaid group medical practice. *Medical Care* 19:2:172-183.

McCord, M.J. (2001): Health care microinsurance - case studies from Uganda, Tanzania, India and Cambodia. *Small Enterprise Development* 12:1:25-38.

McCullagh, P. and Nelder, J.A. (1989): *Generalized linear models*. London.

McIntyre, D., and Owen, P. (1994): *The structure and financing of South African health services: Future options*. Discussion Paper 118. Centre for Health Economics, York.

McIntyre, D., Bloom, G., Doherty, J. and Brijlal, P. (1995): *Health expenditure and finance in South Africa*. Health Systems Trust and World Bank. Durban.

McIntyre, D. (1993): *Health sector financing and expenditure in South Africa*. Health Economics Unit Working Paper No 4, University of Cape Town.

McLean, R.A., Sanders, W.L. and Stroup, W.W. (1991): A unified approach to mixed linear models. *The American Statistician*, 45:54-64.

Medical Scheme Act, Act 72 (1967). The Republic of South Africa, Pretoria.

Medical Scheme Amendment Bill (1989). The Republic of South Africa, Pretoria.

Medical Scheme Amendment Act, Act 55 (2001). The Republic of South Africa, Pretoria. Available online: http://www.polity.org.za/html/govdocs/legislation/2001/act55.htm

Medical Scheme Amendment Act, Act 62 (2002). The Republic of South Africa, Pretoria. Available online: http://www.polity.org.za/pdf/MedicalSchemesAA62.pdf

Mellor, A.C., Blinkhorn, A.S., Hassall, D.C., Holloway, P.J., and Worthington, H.V. (1997): An assessment of capitation in the General Dental Service contract 2. Patterns of treatment provided to regularly attending patients. *British Dental Journal* 182:12:460-464.

Menard, S.W. (2001): *Applied logistic regression analysis.* Thousand Oaks.

Merrill, J., Jackson, C. and Reuter, J. Factors that affect the HMO enrollment decision: a tale of two cities. *Inquiry* 22:2:388-395.

Miller, R.H. and Luft, H.S. (1997): Does managed care lead to better or worse quality of care? *Health Affairs* 16:5:7-25.

Miyazaki, H. (1977): The rat race and internal labour markets. *The Bell Journal of Economics* 9:394-418.

Miriam, D.H. (2003): Indigenous social insurance as an alternative financing mechanism for health care in Ethiopia (the case of eders). *Social Science and Medicine* 56:8:1719-1726.

Mooney G. (1987): What does equity in health mean? *World Health Statistics Quarterly* 40:4:296-303.

Moskowitz, D.S. and Hershberger S.L. (eds.) (2002): *Modeling intraindividual variability with repeated measures data.* Multivariate Applications Series, New Jersey.

Mossialos, E. and Thomson, S.M.S. (2002): Voluntary health insurance in the European Union: a critical assessment. *International Journal of Health Services.* 32:1:19-88.

Nelder, J.A. and Wedderburn, R.W.M. (1972): Generalized linear models. *Journal of the Royal Statistical Society.* Series A, 135:370-384.

n/e/r/a (1993): *Financing health care with particular reference to medicines, Vol. 5: The health care system in Germany.* n/e/r/a - National Economic Health Associates. New York.

n/e/r/a (1996): *Financing health care with particular reference to medicines, Vol. 19: The health care system in South Africa.* n/e/r/a - National Economic Health Associates. New York.

Neudeck, W. and Podczeck, K. (1996): Adverse selection and regulation in health insurance markets. *Journal of Health Economics* 15:4:387-408.

Neuhaus, J.M., Kalbfleisch, J.D. and Hauck, W.W. (1991): A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* 59:25-35.

Neuhaus, J.M. (1992): Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research* 1:249-273.

Newbrander, W., Collins, D.H. and Gilson, L. (2000): *Ensuring equal access to health services: user fee systems and the Poor.* Management Science for Health, Boston.

Newhouse, J.P. and the Insurance Experiment Group (1981): Some interim results for a controlled trial of cost sharing in health insurance. *New England Journal of Medicine* 305:1501-7.

Newhouse, J.P. (1989): Adjusting capitation rates using objective health measures and prior utilisation. *Health Care Financing Review* 10:3:41-54.

Newhouse, J.P. (1993): Risk adjustment for a children's capitation rate. *Health Care Financing Review* 15:1:39-54.

Newhouse, J.P. and the Insurance Experiment Group (1993). *Free-for-all: health insurance, medical costs, and health outcomes: The results of the health insurance experiment.* Cambridge.

Newhouse, J.P. (1996): Reimbursing health plans and health providers: selection versus efficiency in production. *Journal of Economic Literature* 34:1236-1263.

Newhouse, J.P, Buntin, M.J. and Chapman J.D. (1997): Risk adjustment and Medicare: taking a closer look. *Health Affairs* 16:5:26-43.

Newhouse, J.P (1998): Risk adjustment: where are we now. *Inquiry* 35:2:122–131.

Newhouse, J.P., Buntin, M.B. and Chapman, J.D. (1999): *Risk adjustment and Medicare.* Report to the Commonwealth Fund New York. Available online: http://www.cmwf.org/programs/medfutur/newhouse_riskadj_revised_232.asp)

Nyman, J.A. (2003): *The theory of demand for health insurance.* Stanford.

Padoa, F. (1984): Comment on the annual lecture of Prof. Walter A. Weisskopf. *The Geneva Papers on Risk and Insurance* 9:33:366-369.

Pauly, M.V. (1968): The economics of moral hazard: Comment. *American Economic Review* 58:3:531-68.

Pauly, M.V. (1983): More on moral hazard. *Journal of Health Economics.* 2:1:81-6.

Pauly, M. V. (1984): Is cream-skimming a problem for the competitive insurance market? *Journal of Health Economics* 3:1:87–95.

Pauly, M.V. (1986): Taxation, health insurance, and market failure in the medical economy. *Journal of Economic Literature* 24:2:629-675.

Pauly, M.V. and Herring, B. (1999): *Pooling health insurance risks.* Washington DC.

Pauly, M.V. (2002): Why the United States does not have universal health insurance: a public finance and public choice perspective. *Public Finance* Review 30:5:349-365.

Peltzman, S. (1989): The economic theory of regulation after a decade of deregulation *Brooking Papers on Economic Activity: Microeconomics* 1989:1-59.

Popper, K.R. (1945): *The open society and its enemies.* London.

Price, J.R. and Mays, J.M and Trapnell, G.R. (1983): Stability in the Federal Employees Health Benefits Program. *Journal of Health Economics* 2:3:207-23.

Price, J.R. and Mays, J.M. (1985): Biased selection in the Federal Employees Health Benefits Program. *Inquiry* 22:67-77.

Price, M., Khunoae, B. and van den Heever, A. (1996): *Should tax-concession on Medical Aid contributions be removed? A political-economic analysis.* Centre for Health Policy Paper No. 39, University of Witwatersrand, Johannesburg.

Propper, C. (1989): An econometric analysis of the demand for private health insurance in England and Wales. *Applied Economics* 21:777-792.

Propper, C. (1993): Constrained choice sets in the UK demand for private medical insurance. *Journal of Public Economics* 51:287-307.

Propper, C., Rees H. and Green, K. (2001): The demand for private medical insurance in the UK: A cohort analysis. *The Economic Journal* 111:471:180-200.

Purohit, B.C. (2001): Private initiatives and policy options: resent health system experience in India. *Health Policy and Planning* 16:1:87-97.

Rama, P. and McLeod, (2001): *An historical study of trends in medical schemes in South Africa: 1974 to 1999.* Center for Actuarial Research Monograph No.1, University of Cape Town.

Ramaswamy S., Aravas N., Wildasin D.E., Wilson J. D. (1998): Risky local tax bases: risk-pooling vs. rent-capture. *Journal of Public Economics* 69:2:229-247(19).

Ramsay, J.O. (2002): Multilevel modeling of longitudinal and functional data. In: Moskowitz, D.S. and Hershberger S.L. (eds.): *Modeling intraindividual variability with repeated measures data.* New Jersey, Chapter 4.

Rice, N. and Smith, P.C. (2000): Capitation funding in Australia: imperatives and impediments. *Health Care Management Science* 3:77-88.

Rice, N., Dixon P., Lloyd, D.C., Roberts, D. (2000): Derivation of a needs based capitation formula for allocating prescribing budgets to health authorities and primary care groups in England: regression analysis. *British Medical Journal* 320:7230:284-288.

Riley, J.G. (1979): Noncooperative equilibrium and market signalling. *American Economic Review* 69:2:303-307.

Riley, J.G., Tudor, C., Chiang, Y. and Ingber, M. (1996): Health status of Medicare enrolees in HMOs and Fee-for-Service in 1994. *Health Care Financing Review* 22:65-67.

Rothschild, M. and Stiglitz, J.E. (1976): Equilibrium incompetitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economic.* 90:629-649.

Rousseeuw, P.J. and Robust, L.A.M. (1987): *Regression and Outlier Detection.* New York.

Sahai, H. and Ageel, M.I. (2000): *The analysis of variance - Fixed, random and mixed models.* Boston.

Sapelli; C. and Torche, A. (2001): The mandatory health insurance system in Chile: Explaining the choice between public and private insurance. *International Journal of Health Care Finance and Economics* 1:2:97-110.

Sapelli, C. and Vial, B. (2003): Self-selection and moral hazard in Chilean health insurance. *Journal of Health Economics* 22:3:459-476.

Savage, E. and Wright, D. (2003): Moral hazard and adverse selection in Australian private hospitals: 1989-1990. *Journal of Health Economics* 22:3:337-359.

Scheffler, R.M. and Rossiter, L.F. (eds.) (1985): *Advances in health economics and health services research* Vol 6. Greenwich.

Schokkaert, E., Dhaene, G. and Van de Voorde, E, C. (1998): Risk-adjustment and the trade-off between efficiency and risk selection. *Health Economics* 7:465–480.

Schokkaert, E. and van de Voorde, C. (2000): Risk adjustment and the fear of markets: the case of Belgium. *Health Care Management Science* 2:121-130.

Schumpeter, J.A. (1921): *The theory of economic development.* Oxford.

Scitovsky, A.A., McCall, N. and Benham, L. (1978): Factors affecting the choice between two prepaid plans. *Medical Care* 16:8:660-675.

Shmueli, A., Shamai, N., Levi, Y., and Abraham, M. (1998): In search of a natural capitation formula: the Israeli experience. In: Chinitz, D. and Cohen, J. (eds.): *Governments and health systems--Implications of differing involvements.* New York, pp.201-208.

Shmueli, A. (1999): Inferring capitation rates from aggregated health plan costs. *Health Economics Letters* 8:547-552.

Shmueli, A., Chernichovsky, D. and Zmora, I. (2003): Risk adjustment and risk sharing: the Israeli experience. *Health Policy* 65:1:37-48 Special Issue.

Smith, P., (1998): The estimation, application, and experience of the new capitation formula in the United Kingdom. In: Chinitz, D. and Cohen, J. (eds.): *Governments and health systems—Implications of differing involvements.* New York, pp.209-224.

Soderlund, N., Schierout, G. and van den Heever, A. (1998): *Private health care in South Africa.* Report to Health Systems Trust by the Centre for Health Policy. Johannesburg.

Soderlund, N. and Hansl, B. (1999): *Health insurance in South Africa - the impact of de-regulation on equity and efficiency.* Centre for Health Policy Monograph No. 57. University of Witwatersrand, Johannesburg.

Soderlund, N. and Hansl, B. (2000): Health insurance in South Africa: An empirical analysis of trends in risk-pooling and efficiency following deregulation. *Health Policy and Planning* 15:4:378-385.

South African Republic (1994): *Report of the Commission of Inquiry into the manner of providing for medical expenses.* (Melamet Commission Report). Pretoria.

STATA (1999): *Reference Manual Release 6.* STATA Press.

Stewart, J. (1994): The welfare implications of moral hazard and adverse selection in competitive insurance markets. *Economic Inquiry* 32:2:193-208

Steyn, N. (1996): *How much are we ACTUALLY Spending on Health? An urgent call for consolidated information.* Pretoria.

Stiratelli, R., Laird, N.M. and Ware, J.H. (1984): Random-effects models for serial observations with binary response. *Biometrics* 40:961-971.

Tversky, A. and Kahnemann, D. (1982): Availability: A heuristic for judging frequency and probability. In: Kahnemann, D., Slovic, P. and Tversky, A. (eds.): *Judgement and uncertainty: Heuristic and biases.* Cambridge, pp.163-178.

UNECA (2003): *Economic Report on Africa 2002: Tracking Performance and Progress.* UN Economic Commission for Africa. Addis Ababa.

Van Barneveld, E., van Vliet, R. and van de Ven, W., (1996). Mandatory high risk-pooling: an approach to reducing incentives for cream-skimming. *Inquiry* 33:133-143.

Van Barneveld, E., Lamers, L.M., van Vliet, W. R. and van de Ven, W. (1998): Mandatory pooling as a supplement to risk-adjusted capitation payments in a competitive health insurance market. *Social Science and Medicine* 47:2:223-232.

Van Barneveld, E., Lamers, L., van Vliet, R. and van de Ven, W., (2000). Ignoring small predictable profits and losses: a new approach for measuring incentives for cream-skimming. *Health Care Management Science* 3:131-140.

Van Barneveld, E., Lamers, L., van Vliet, R. and van de Ven, W. (2001): Risk sharing as a supplement to imperfect capitation: a tradeoff between selection and efficiency. *Journal of Health economics* 20:147-168.

Van den Heever, A.M. (1997): Health Care Financing. In: *South African Health Review 1997.* Health Systems Trust. Durban.

Van den Heever, A.M. (1998): Private sector health reform in South Africa. *Health Economics* 7:4:281-289.

Van de Ven, W. and van Vliet, R. (1992): How can we prevent cream skimming in a competitive health insurance market? The great challenge for the 1990s. In: Zweifel, P. and Frech III, H.E. (eds.): *Health economics worldwide.* Dordrecht, pp. 23-46.

Van de Ven, W. and van Vliet, R. (1994): Risk-adjusted capitation: recent experiences with The Netherlands. *Health Affairs* 13:5:120-135.

Van de Ven, W., van Vliet, R., van Barneveld, E., Lamers, L. (1998): Towards a risk-adjustment mechanism in a competitive health insurance market. In: Chinitz, D. and Cohen, J. (eds.): *Governments and health systems--Implications of differing involvements.* New York, pp. 179-184.

Van de Ven, W., van Vliet, R., Schut, F. and van Barneveld, E. (2000): Access to coverage for high-risks in a competitive individual health insurance market: via premium rate restrictions or risk-adjusted premium subsidies. *Journal of Health Economics* 19:311-339.

Van de Ven, W. and Ellis, R., (2000): Risk adjustment in competitive health plans. In: Culyer, A.J. and Newhouse, J. (eds.): *Handbook of health economics.* Amsterdam, Chapter 17.

Van de Ven, W., Beck, K., Buchner, B. et al. (2003): Risk adjustment and risk selection on the sickness fund insurance market in five European countries. *Health Policy* 65:1:75-98.

Van Vliet, R. (1992): Predictability of individual health care expenditures. *Journal of Risk and Insurance* 59:3.443–461.

Van Vliet, R. and van de Ven, W. (1992). Towards a capitation formula for competing health insurers. An empirical analysis. *Social Science and Medicine* 34:9:1035–1048.

Van Vliet, R.C., van de Ven, W.P. (1993): Capitation payments based on prior hospitalizations. *Health Economics* 2:2:177-88.

Van Vliet, R. (2000): A statistical analysis of mandatory pooling across health insurers. *Journal of Risk & Insurance* 67:2:197-217.

Wallace, D. and Green S.B. (2002) Analysis of repeated measures design with linear mixed models. In: Moskowitz, D.S. and Hershberger S.L. (eds.): *Modeling intraindividual variability with repeated measures data.* Multivariate Applications Series, New Jersey, pp.103-134.

Wallack, S.S., Tompkins, C.P. and Gruenberg, L. (1988): A plan for rewarding efficient HMOs. *Health Affairs* 7:80-97.

Ware, J.H. (1985): Linear models for the analysis of longitudinal studies. *The American Statistician* 39-95-101.

Weiner, J.P, et al. (1995): Risk-adjusted Medicare capitation rates using ambulatory and inpatient diagnostics. *Health Care Financing Review* 17:3:77-101.

Weisskopf, W.A. (1984): Reflections on uncertainty in economics. Seventh annual lecture of the Geneva Association (International Association for the Study of Insurance economics). *The Geneva Papers on Risk and Insurance* 9:33:335-360.

Welch, W.P. (1989): Restructuring the Federal Employees Health Benefits Program: the private sector option. *Inquiry* 26:3:321-34.

Weng, K.Z. (1999): Risk-pooling over demand uncertainty in the presence of product modularity - Continuous univariate distributions. *International Journal of Production Economics* 62:1:75-85(11).

Whitehead, M. (1990): *The concepts and principles of equity in health.* Report for the WHO Regional Office for Europe, Copenhagen.

Whitehead, M. (1992): The concepts and principles of equity in health. *International Journal for Health Services* 22:3:429-445.

Whithford, A. and McGrath, M. (1994): *Income distribution, inequality and poverty in South Africa: Preliminary findings from the Project for Statistics in Living Standards and Development.* SALDRU Conference on Poverty. South Africa.

Wiesmann, D. and Juettig, J, (2000): The emerging movement of community based health insurance in sub-Saharan Africa: experiences and lessons learned. *Afrika Spectrum* 35:2:193-210.

Wilson, C.A. (1977): A model of insurance markets with incomplete information. *Journal of Economic Theory* 16:2:167-207.

Wilson, C.A. (1979): Equilibrium and adverse selection. *American Economic Review* 69:2:313-317.

Wilson, C. A. (1980): The nature of equilibrium in markets with adverse selection. *Bell Journal of Economics* 11:1:108-130.

Wolfe, J.R. and Goddeeris, J.H. (1991): Adverse selection, moral hazard, and wealth effects in the medigap insurance market. *Journal of Health Economics* 10:4:433-459.

World Bank (1993): *Investing in Healt.:* World Development Report. New York.

World Bank (1994): *Better Health in Africa.* Washington.

World Bank (1988): *Adjustment Lending: An Evaluation of Ten Years of Experience* Washington.

World Health Organisation (1993): *Planning and implementing health insurance in developing countries: Guidelines and case studies.* Macroeconomics, Health and Development Series Nr. 7, Office of International Cooperation. Geneva.

Wrightson, W., Genuardi J. and Stephens, S. (1987): Demographic and utilization characteristics of HMO disenrollees. *GHAA Journal* 23-42.

Wubben, E.F.M. (1993): Markets, uncertainty and decision-making: a history of the introduction of uncertainty into economics. Amsterdam.

Yu, W., Ellis, R.P. and Ash, A. (2003): Risk selection in Massachusettes State Health Insurance Program. Unpublished, Running title: Risk selection in health insurance. Available online: http://econ.bu.edu/Ellis/Papers/MassGIC.pdf

Zeger, S. and Liang, K.Y. (1986): Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42:121-130.

Zeger, S. and Liang, K.Y. and Albert, P.S. (1988): Models for longitudinal data: A generalized estimation equation approach. Biometrics 44:1049-1060.

Zeger, S. and Liang, K.Y. (1992): An overview of methods for the analysis of longitudinal data. Statistics in Medicine 11:1825-1839.

Zweifel, P. and Frech III, H.E. (eds.) (1992): Health economics worldwide. Dordrecht

Zweifel, P. and Breyer, F. (1997): Health economics. New York.

Zweifel, P. (ed.) (1998): Health, the medical profession, and regulation. Dordrecht.

Appendix 1

Case study background: Republic of South Africa

The following background for the case study of an unregulated health insurance market concerns facts about the Republic of South Africa. In addition to the presentation of a more general introduction to this African country, this section will also emphasise major economic and political developments that influenced the South African health care system between 1995 to 1998 - the time frame of the thesis' case study and empirical analysis. The aim is to put the data analysis in the later part of the thesis into the context of the events that occurred during the years of data collection.

The Republic of South Africa occupies the southernmost part of the Sub-Saharan African continent. It is divided into nine provinces, each with its own Legislature, Premier and Ministers. The provinces are the Western Cape, the Eastern Cape, KwaZulu-Natal, the Northern Cape, Free State, North-West, Gauteng, Mpumalanga, and the Northern Province. Before 1993 South Africa was divided into four 'independent states' (Transkei, Bophuthatswana, Venda, and Ciskei), six 'self-governing territories' (KwaZulu, KaNgwane, QwaQwa, Lebowa, Gazankulu, and KwaNdebele), and four provinces of 'White South Africa' (Cape, Natal, Orange Free State, and Transvaal (see figure A1-1).

South Africa's population accounts for 40.58 million people made up of Black Africans (76.7%), White Africans (10.9%), Coloured Africans (8.9%), and Indian/Asian Africans (2.6%). More than half of the population lives in urban areas. These numbers are the result of the first official census in post-apartheid South Africa in October 1996[167] (Government of South Africa 2003). The country's population estimates in the year 2001 stood at 44.6 million, of which 52% were women. The annual population growth rate is 2.1% (UNECA 2003).

South Africa's history is tied to the apartheid system that was established after the National Party (NP) won the elections in 1948. In most respects, apartheid was a more systematic and brutal continuation of the segregationist policies of previous governments. However, a new concern with racial purity was apparent in laws prohibiting interracial sex and in provisions for population registration requiring that every South African be assigned to one discrete racial category or another. The apartheid state also created a

[167]The second democratic census was held in October 2001. First census results are expected in Summer 2003.

permanent White political majority by eliminating the voting rights of all Black Africans. 'Homelands' for Africans (and coloured people) were created where alternative political provisions could be made. This included the enforcement of total segregation, so that every town was carved into separate 'group areas', sorting people by racial categories as shown in their identity books and entered in a national register.

South Africa's first democratic election took place in April 1994, and led to a Government of National Unity (GNU). A new constitution, adopted by Parliament in May 1996, came into force in February 1997 after certification by the Constitutional Court. Soon afterwards, the NP withdrew from the GNU, leaving only the IFP (Inkatha Freedom Party) to partner with the African National Congress (ANC) in the Cabinet. South Africa is still facing enormous challenges at all society levels in regard to its former segmentation along racial lines.

South Africa's economy faced serious problems after the abolition of apartheid in the middle 1990's.[168] Two main issues have dominated the economic performance of South Africa since 1995. First, the Rand went through three currency crises, 1996, 1998, and 2001, and fell steadily against other currencies.[169] Second, in 1996, the Government announced its macroeconomic stabilisation strategy, which incorporated its priorities for Growth, Employment and Redistribution (GEAR), based on the International Monetary Fund (IMF) austerity measures.[170] Economic growth for the last two years (2001/2002 and 2002/2003) has been disappointing. While growth in total real gross domestic product (GDP) averaged 2.6% (0.5% per capita) between 1996 and 2001 (UNECA 2003), the weakening of the international economy was reflected in a slowdown of growth during 2002. South Africa, like other developing economies, is highly susceptible to trends in the economies of its major trading partners (the United States, United Kingdom, Europe and the Far East) and regional political instabilities still negatively affect investor perceptions.

[168] Apartheid also included the retention of economic power in white hands, by tightening the job colour bar and directing skilled blacks into their own areas. This was linked to a policy of industrial decentralisation, so that centres of industry could be set up on the borders of homelands, to which black and white employees could travel from opposite sides without infringing group areas (McIntyre et al. 1995).

[169] Also, the price of gold – on which the extensive domestic mining industry relies - fell on the international market in concert with the Rand. This made major reductions in mine labour necessary just when an increase employment in the industrial sector was urgent. At the same time, the good prospects for South Africa's export trade offered by a cheaper Rand were weakened by the sudden collapse of Far Eastern markets, to which South Africa looked for expansion (Government of South Africa 1999).

[170] In particular, the vision of a society in which sound health, education and other services are available to all was stated with commitments to financing low-income housing, a national school feeding programme, water and sanitation projects, free primary health care services, and enhanced small business support.

Despite the fact that South Africa has the most advanced economy on the African continent, and has been highly commended for its successful macro-economic policies, wide disparities of wealth remain, with obvious implications for broader socio-political policy directions. Although South Africa is classified as a middle-income country great inequalities in the distribution of income are reflected in the following numbers: 51% of the income belonged to the richest 10% of the households and only 4% to the poorest 40% (World Bank 1994). The Gini-coefficient of income concentration for 1994 was 0.65 (Whiteford and McGrath 1994), more than 20% higher than for other middle-income countries. This strong economic segmentation of South Africa's society can be also seen in the health sector.

Under the apartheid system health care financing and health care delivery was grossly inequitable.[171] Naturally, it was one of the highest priorities of the new democratic South African government to change this for the better. Three constitutional imperatives are the guidelines for today's health legislation and policy: (1) realise the right of access of health care to all, in particular, (2) promote children's right to basic health care and (3) ensure that nobody is refused emergency medical care[172] (HST 2003). The *White Paper on the Transformation of the South African Health System* (DoH 1997b) set out a plan for decentralising health-care delivery as well as re-orientating services to focus on primary health care. Its objectives were to unify the fragmented health services at all levels into a comprehensive and integrated national health system. However, the envisioned *National Health Bill*, which has already taken more than seven years to draft, is not expected to pass the Parliament until the end of 2003 (HST 2003). The major goal of the Bill is to create a National Health System that includes public, private, and non-governmental providers and to coordinate services equitably and efficiently[173] (DoH 1996).

---

[171] Under the apartheid system health care was separated along racial lines and health care financing according to the racially separated geographic regions (McIntyre et al. 1995). White Africans had not only higher income, the governments spent more per person on services in 'White' areas. Because the White population was concentrated in urban areas health care facilities here were well-developed and comparable to developed countries. However, half of the South African population, mostly Black Africans, lived in peripheral areas at the household subsistence level with inadequate health care provision (McIntyre et al. 1995). White, high-income earners were often members of insurance plans - the medical schemes - and obtained their medical care from the private sector. Some Non-White, low to middle-income earners (one third of the population) living in poor urban areas had some access to low-cost medical schemes (as a result of their strong unionisation since 1979). These low-cost schemes are usually attached to employer's companies or industrial sectors (Soderlund and Hansl 1999). However, half of the South African population, mostly Black Africans, lived in peripheral areas at the household subsistence level with inadequate health care provision (McIntyre et al. 1995).

[172] South African Constitution, articles 27 (1), (2) and (3) and 28 (1).

[173] It will also establish a legal framework for the district health system and the regulation of public and private facilities. It will detail the rights and duties of health-care users and providers, set up a complaints

Two trends, in public and private health care financing could be observed since the early 1990's. Falling per capita public health care financing followed a period of substantial growth in public health care expenditures, between 1992 and 1998.[174] Private health care expenditure rose while the number of people with access to private care fell, putting an additional burden on the overstretched resources of the public sector (HST 2003).[175] Together with limited economic growth in the late 1990's and persistent public sector inefficiencies this interrupted the initial trend of increasing access and equity in South Africa's health care, starting after the first democratic election in 1994.

The term "equity"[176] is commonly found in publications regarding the South African health care sector and has been a goal of many South African health policies. In international health policy debates of the last decade efficiency-driven perspectives started to dominate. For example, there has been a rapid development of approaches aimed at cost-effective rationing of scarce resources for health care. Some progress has been made

mechanism, address access to and confidentiality of information, and provide mechanisms for national and provincial health authorities to consult with all stake-holders.

[174] According to the latest data available for 1998-1999 the government accounts for 44% of South Africa's health care financing sources, 39% is contributed by private households, and 17% by private employers. A small amount of 0.1% comes from donors and NGOs. (HST 2003).

[175] Even though more than 60% of all health care expenditure go to the private sector it serves less than 20% of the population (DoH 2001). Moreover, private sector medical scheme coverage declined as a proportion of the total population according to the latest 1998-1999 data. This probably reflects rising unaffordability of medical scheme membership as costs escalated in the private sector (HST 2003).

[176] Global health professionals have struggled with a definition for equity for some time. Mooney (1987) went even that far to conclude that equity is a value laden concept which has no uniquely correct definition (after discussing five theories of equity in distribution of health resources: 1. A theory of maximum - Rawl's theory modified to include health care institutions providing opportunity as the social good; 2. Altruism as a basis for equity - Titmuss' Kantian view of national responsibility to provide equitable service delivery altruistically or equal access; 3. A fair share theory of distribution - Margolis' process utility theory of doing one's fair share or equality of access for equal need; 4. Commitment to equity - Sen's focus on sympathy and commitment to another's ill health status and access; and 5. Equity as externality - Culyer's health care consumption where government determines the merit good or extent of consumption). Other authors were more comfortable providing definitions of equity in the health literature: (1) Whitehead (1990): "Equity in health implies that ideally everyone should have a fair opportunity to attain their full health potential and, more pragmatically, that no-one should be disadvantaged from achieving this potential, if it can be avoided." (2) Berman et al. (1989): "A common definition of equity in the public health literature is that the primary determinant in the use of services should be the need for them. Other factors such as income, race, location of residence and so forth should not play an important role in selecting who receives care and who does not." (3) Newbrander et al. (1995): "Equity means equal opportunity of use of health services for equal need", and (4) Daniels (1982): "Access to health care is equitable if and only if there are no information barriers, financial barriers, or supply anomalies that prevent access to a reasonable or decent basic minimum of health care services.". The consensus definition arrived at is from the World Health Organisation (WHO), developed by Margaret Whitehead (1990). Acknowledging that inequity has a moral and ethical dimension it refers to differences in health, which are not only unnecessary and avoidable, but in addition are considered unfair and unjust in the context of a specific society. Thus equity in health was defined by the WHO as minimising avoidable disparities in health and its determinants – including but not limited to health care – between groups of people who have different levels of underlying social attributes (income, gender, ethnicity, geography, etc).

in reducing mortality and morbidity and in reducing inequalities in health and access to health care. However, many avoidable inequalities in health persist.

Equity was an essential feature of the redistributive policies in post-apartheid South Africa, aspiring to reduce the significant levels of inequality.[177] The discussion of equity in the political arena of South Africa includes most frequently three views of equity, i.e. focus on the health of the most vulnerable, inclusion and narrowing gaps between the poorest and the richest segments of the population in health care access[178] and consumption. The manner in which these equity-oriented policies were (and were not) implemented over the past few decades and the factors that influenced this require further analysis. However, high levels of aggregate ill health - exacerbated by HIV/AIDS - and persistent inequalities in ill health, mortality and access to health care indicate that distributional issues remain inadequately addressed. There exists a widening gap between necessary and actual public allocations for health, and between the private and public health sector. Resources are concentrated either in central hospitals or in urban private care, confronting disadvantaged communities with spiralling health care costs. Those with the greatest health needs are often those with the least ability to access good quality health services. One of the strongest determinants for the consumption of good quality health care remains the (private) health insurance status of a person (Hansl and Soderlund 1999). Private health insurance only guarantees South Africans access to good quality health

[177] It is important to clarify the understanding of equity and equality. Although these terms have different meanings, they are frequently used interchangeably. Equity adheres to predetermined norms or standards, which are considered fair or just when describing gaps, differences or disparities. These norms or standards vary from place to place, from time to time and from one community to the other. Due to the difficulties in setting these norms or standards – frequently influenced by values and judgments – the notion of inequality is more frequently used. Contrary to equity, equality does not take into account whether the existing gaps, differences or disparities are fair or just. In other words, inequity is unjust or unfair inequality.

[178] There is of course some ambiguity regarding the term access. Having access to health care does not automatically lead to utilisation. Transforming potential access into realised access or utilisation is depending on a dynamic interaction with demand or felt need and various enabling factors such as ability and willingness to pay, travel time, quality of care, etc.

services. Thus in South Africa obtaining health insurance is widely discussed in terms of equitable access to good quality health care.

Appendix 2

The basic logit model and its assumptions

Logit models are able to approach empirical problems where it is of interest to estimate the effect of the covariates on a dichotomous outcome with odds ratios. For example, this thesis seeks to estimate the effect of selected variables on the likelihood that an event occurs, i.e. if households are more likely to join, leave, or switch insurance coverage rather than not. The thesis considers the analysis of a panel of N households, having recorded histories that indicate the presence or absence of an event in each T equally spaced time period. These types of statistical models, which have an endogenous variable that only takes discrete values, are known as categorical, qualitative-dependent or qualitative-choice models. The most widely used discrete-response models are the linear probability model, the probit model, and the logit model.

Logit analysis or logistic regression employs the logit model as one type of discrete-response model. In most cases, like the empirical analysis in this thesis, the dependant variable y only assumes two values - 1 if an event occurs and 0 if it does not. These models are referred to as models with dichotomous dependant variables. For example, applying the analytical problem of this thesis, a logit analysis here would comprise a dependent variable y with the value 1 if an individual was a joiner, leaver or switcher and zero if in the control group (for i= 1, ...N households). Suppose now that y takes the value 1 with the probability p, and y takes the value 0 with the probability (1-p). Then one can formulate the expected value of y, Ey as the probability that the event will occur as the following:

$$Ey = 1p + 0(1-p) = p \qquad \textit{(Equation A2-1: Event probability)}$$

If we assume that the probability is a function of a vector of independent explanatory variables, $\mathbf{x}_i$ and of a vector of unknown parameter $\mathbf{\beta}$ the general discrete model is the following:

$$\text{Prob}(y_i = 1) = F(\mathbf{\beta}'\mathbf{x}_i) = E(y_i \mid \mathbf{x}_i) \quad \textit{(Equation A2-2: Discrete model probability function)}$$

As mentioned above the three most applied functional forms for F are the linear probability model, the probit model, and the logit model (Hisao 1986). With the linear probability model it is possible to have estimated probabilities outside the 0-1 range because it does not constrain $\mathbf{\beta}'\mathbf{x}_i$ values to lie between 0 and 1. The logit and the probit

model limit the probability between 0 and 1. The probability functions for probit models use the standard normal distribution and for logit models use the logistic distribution. The logit model is the following:

$$F(w) = \frac{e^w}{1+e^w} \qquad \textit{(Equation A2-3: Logit model)}$$

In the simple linear regression model we assume that an observation of a random outcome variable y can be expressed as a function of $\mathbf{x}$, $y = \beta'\mathbf{x} + v$. For the error term v the most common assumption is that it follows a normal distribution with a mean of zero and some variance that is constant across the levels of the independent variable. This is different for cases with a dichotomous outcome variable. The logit models correspond to the cumulative distribution of v being logistic (Hisao 1986).

Logistic regression is used to predict the likelihood that an individual case will fall into one of two possible groups or classifications, similar to discriminant analysis - another tool for examining binary, non-interval data. However, logistic regression is not as restrictive as discriminant analysis in its assumptions about the characteristics of the data used. This makes logistic regression analysis preferable over two-group discriminant analysis (Hair et al. 1998). Multiple linear regression and discriminant analysis share fundamental assumptions about normality, homoskedasticity and linearity.[179]

Logistic regression is, in general, less sensitive to underlying assumptions than discriminant analysis. In particular, when the assumption of multivariate normality is violated, the p values derived from logistic regression are still accurate, whereas the p values derived from discriminant analysis are marginally biased.[180] Likewise, violations of the assumption of homogeneous covariance may artificially raise the p values in discriminant analysis but not logistic regression.[181] Logistic regression does not require linear relationships between the independents and the dependent, but it does assume a linear relationship between the logit of the independents and the dependent. Finally, dichotomous predictor variables are more applicable to logistic regression than

[179] Normality means that samples of the independent variables have to be normally distributed. The errors are normally distributed. Homoskedasticity assumes that the populations of the groups being compared have equal covariance (dispersion) matrices. The variance of the errors is constant, i.e. homogeneity of variance. Linearity requires linearity of relationships between variables. The population means of the dependent variables at each level of the independent variable can be arranged on a straight line.

[180] First, however, the distortions to the p values in discriminant analysis are marginal when the sample size is large. Second, both techniques are undermined by violations of normality when the sample size is small. Third, severe violations will also undermine the power and efficacy of logistic regression, despite the accurate p values.

[181] This is not a problem for discriminant analysis when the sample size of each group is similar.

discriminant analysis.[182] Other conditions, such as multicolinearity[183], are also significant for logistic regression. Since the assumptions for multiple regression differ substantially from logistic regression a comprehensive list of these assumptions for logistic regression is provided here (based on Menard 2001).

First of all it has to be ensured that the model is correctly specified which requires:

- **Inclusion of all relevant variables in the model.** If relevant variables are omitted, the common variance they share with included variables may be wrongly attributed to those variables, or the error term may be inflated.
- **Exclusion of all irrelevant variables.** If causally irrelevant variables are included in the model, the common variance they share with included variables may be wrongly attributed to the irrelevant variables. The more the correlation of the irrelevant variable(s) with other independents, the greater the standard errors of the regression coefficients for these independents.
- **Low error in the explanatory variables.** Ideally low measurement error and no missing cases are assumed.

These three assumptions are crucial to consider before and during the model building process. Other conditions for logistic regression are:

- **Cases or error terms are assumed independent.** Violations with serious effects occur, for instance, in correlated samples. In before-after studies, panel data or time-series data subjects cannot always provide multiple observations at different time points. However, specific methods are available to adapt logistic models to handle non-independent (or correlated) data.
- **Large samples.** Logistic regression uses maximum likelihood estimation (MLE) rather than ordinary least squares (OLS) to derive parameters. MLE relies on large-sample asymptotic normality, which means that reliability of estimates decline when there are few cases for each observed combination of X variables.
- **Independent variables are not linear functions of each other.** To the extent that one independent is a linear function of another independent, the problem of multicollinearity will occur in logistic regression. As the independents increase in correlation with each other, the standard errors of the logit (effect) coefficients will

[182] Dichotomous predictor variables should not be subjected to discriminant analysis, although Monte Carlo studies suggest that this issue is trivial.

[183] The assumption of non-multicolinearity means that the independent variables are not linear combinations of each other. Perfect multicolinearity makes estimation impossible, while strong multicolinearity makes estimates imprecise.

become inflated. Multicollinearity does not change the estimates of the coefficients, only their reliability. Nowadays, software (like STATA) automatically accounts for multicolinearity and excludes detected variables from the estimated models.

- **Linearity.** Logistic regression does not require linear relationships between the independents and the dependent, but it does assume a linear relationship between the logit of the independents and the dependent.
- **Additivity.** Logistic regression does not account for interaction effects except when interaction terms (usually products of standardized independents) are created as additional variables in the analysis.
- **Expected dispersion.** In logistic regression the expected variance of the dependent can be compared to the observed variance, and discrepancies may be considered under- or over-dispersion. If there is moderate discrepancy, standard errors will be over-optimistic and one should use adjusted standard error. Adjusted standard error will make the confidence intervals wider.

The test for the last set of assumption should be part of each model building process. For, example, possible multicolinearity should already be considered during the variable selection in the model building process.

Appendix 3

Outliers and missing values

Any multivariate model aims to identify complex relationships, but influencing factors, like outliers, missing values, and violated analysis assumptions can compound across several variables and distort the statistical results. Besides a clear understanding of the assumptions for the specific multivariate technique used, a careful data examination before any analysis is necessary. This will ensure that the statistical requirements for the multivariate analysis are met. This appendix applies diagnostic techniques in order to assess the influence of outliers and missing values.

Outliers are observations characterised by distinct differences as compared to the majority of the sample. On one hand they can provide additional information about unrecognised population characteristics or, on the other hand, they can be non-representative of the population and thus bias the statistical analysis. Before categorically excluding outliers from any analysis they should be evaluated in regard to their potential information value and in regard to the objectives and the context of the study.

Outliers may arise due to procedural factors like data entry errors or coding mistakes. Usually, however, outliers represent an extraordinary event that can either be explained by the researcher or not. If the researcher decides that there is a reasonable explanation for the uniqueness of that observation the outlier is included in the analysis otherwise it is often excluded. In addition to these types of outliers there are others that are unique in their combinations of values across several variables and less in regard to values outside the ordinary range. These observations are typically retained because they might represent a valid population segment. For a more thorough discussion on this topic see Barnett and Lewis (1984) or Rousseeuw and Robust (1987).

For the thesis, data outliers resulting from procedural data entry and coding errors or outliers with values that could be not explained were excluded. During the early data cleaning process a very small number of outliers resulting from data entry/coding errors were excluded. For example there were fewer than ten cases where a gender was miscoded. Instead of the original F for female and M for male, two other letters of the alphabet appeared. Outliers in the form of values that could not be explained by the researcher appeared in the South African data set in the form of 'false negative values'. They are referred to as false negative values since it would be technically impossible for these variables to have a negative value. Thus it can be assumed that negative values

generally emerged from entry mistakes. False negative values were very rare and they are summarised in table A3-1 below. The co-payment variable had most false negative values, 0.2% of all cases. All false negative values were excluded during the data cleaning process.

Table A3-1: Outliers – false negative values in original data

Variable	Number of observations	Negative values	
		Number	Percent
age	1,009,272	1	0
gender (female)	1,011,735	0	0
dependents	1,011,735	0	0
income	1,011,735	0	0
race	1,011,735	0	0
tac (claims)	1,011,695	179	0.02
tco (contributions)	1,011,735	0	0
co2 (co-payments)	1,011,735	1,999	0.20
gp (GP benefits)	958,663	10	0
pub (pub.-hospital benefits)	944,991	69	0.01
spe (specialist benefits)	958,636	148	0.01
prv (priv. Hospital benefits)	954,483	168	0.02
den (dental benefits)	958,453	20	0
opt (optical benefits)	914,708	28	0
par (paramedical benefits)	928,476	51	0.01

Note: This is based on the original data of N=1,011,735 observations over n=353,458 households.

For the remainder of the data several diagnostic techniques were applied in order to identify less obvious outliers. The univariate detection of outliers included the evaluation of the histogram of variable observations. However, not all results are shown here. The histograms of the two variables age and optical benefits were chosen as examples for illustration purposes. Next page figures A3-1 and A3-2 display the histograms for each variable. For the age variable the histogram shows no observations that lie significantly outside the bulk of observations. There are very few observations at the tail ends of the distribution, for ages below 10 and above 90. But since that is a part of a normal age distribution they cannot be classified as outliers. The histogram for the optical benefits shows some very large observations that could be classified as outliers. However, even though a small number of observations, particularly for the later continuous benefit variable, occurred in the outer range of observations, there were none that could not be explained or were outright non-normal.

Figure A3-1: Univariate outlier detection with histogram for variable age

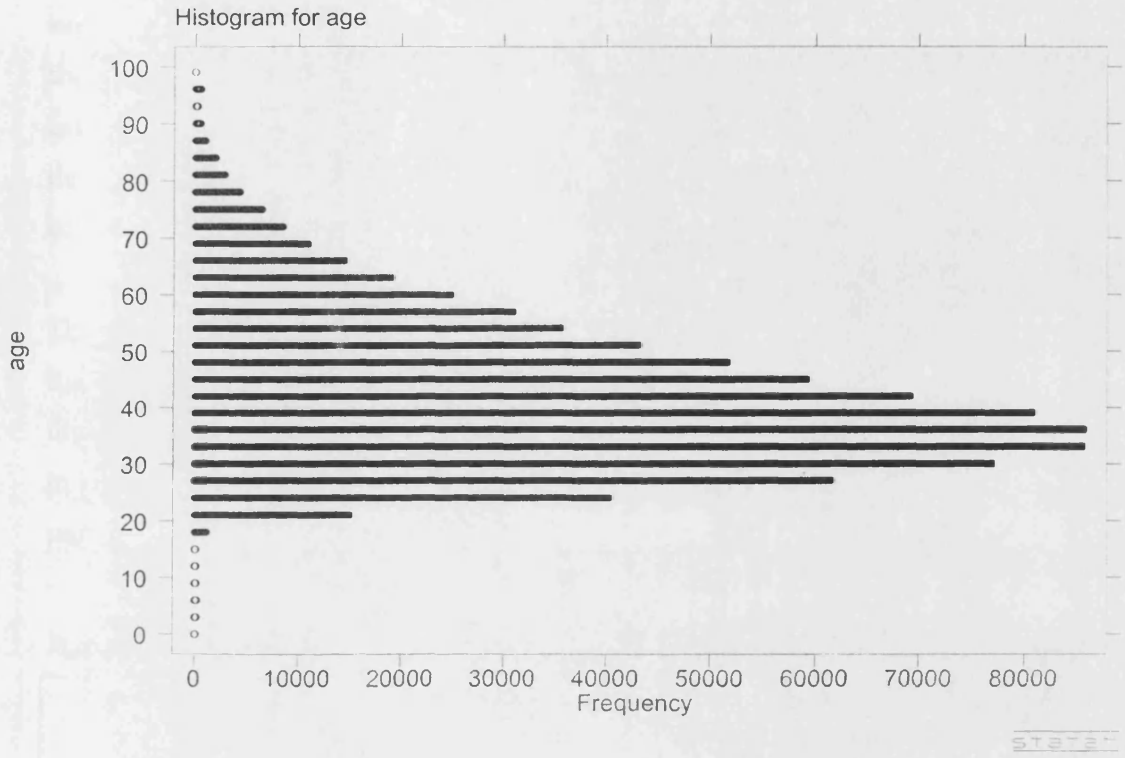
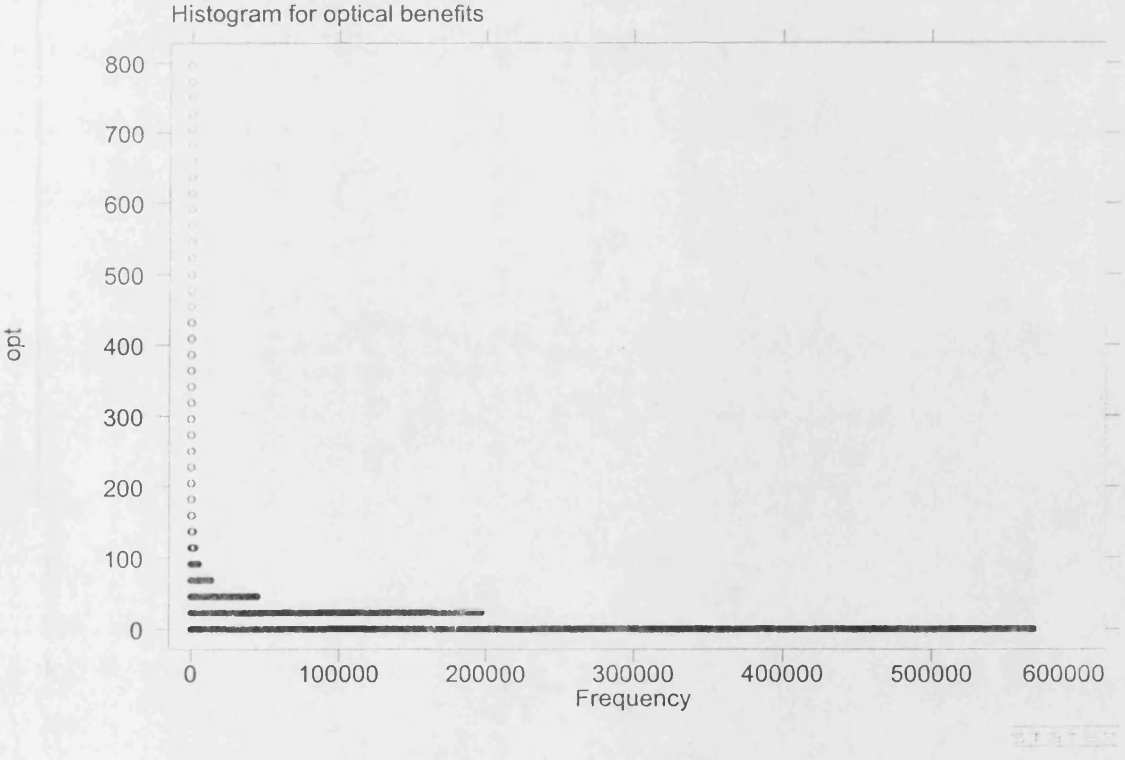


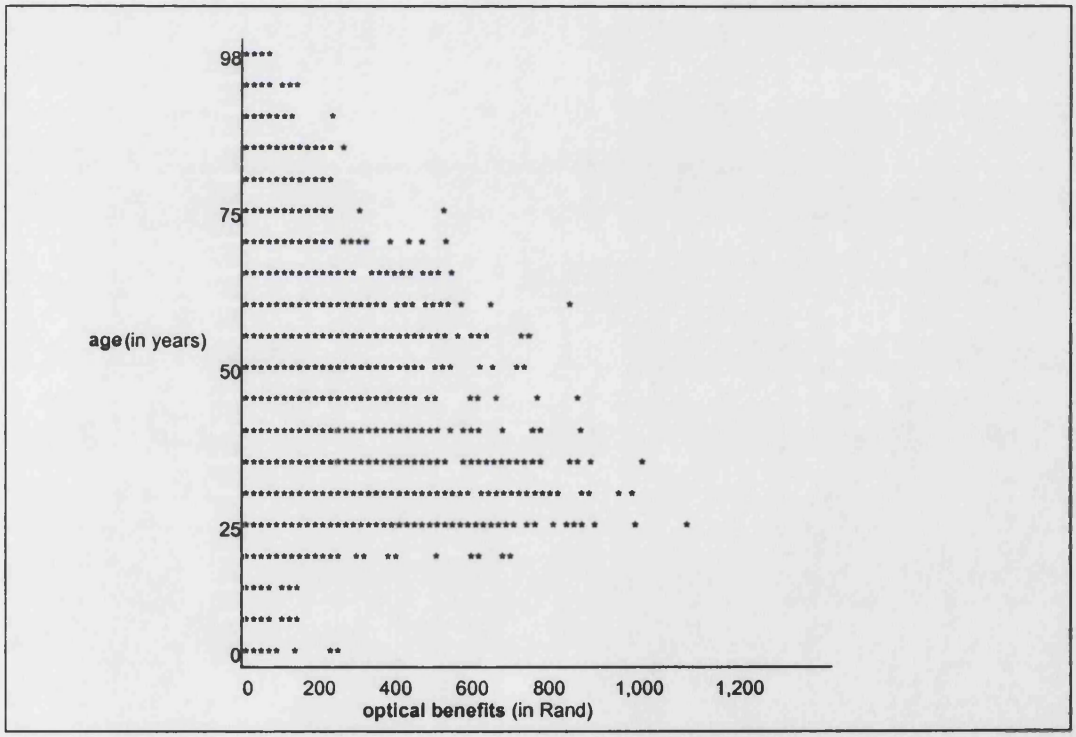
Figure A3-2: Univariate outlier detection with histogram for variable optical benefits



Most of the data variables represent utilisation benefits for different medical services. These medical benefits typically have irregular and extreme patterns because they are due to unexpected, unique events. Also, since it is the objective of this study to investigate the inclusion or exclusion of households from insurance due to unusual socio-demographic or utilisation profiles exclusion of observations with these characteristics could be not justified.

The bivariate outlier detection assessed pairs of variables through a scatterplot. This joint assessment of variables through a scatterplot can help to identify isolated points that fall outside the range of observations as potential outlier cases. Here, serving as an illustration, the scatterplots for the two variables of age and optical benefits are presented in figure A3-3. In the scatterplot it is to be seen that there are no cases that can be noted as particularly isolated points and therefore potentially classified as outliers.

Figure A3-3: Bivariate outliers detection with scatterplot for age and optical benefits



Even if the bivariate outlier detection would have revealed cases that could be potential candidates for an outlier classification, it seems unjustified to exclude them from the following data analysis. All potential outliers represent unique observations that are the result of possibly extreme but explainable events. Also, it is assumed that these

observations represent specific population members, who carry valuable information regarding the influence of these unusual values on their insurance status.[184]

It can be summarised that there is, first, not a large number of outliers. Second, for these outliers no demonstrable proof can be provided that they are truly deviant and do not in fact represent a valid segment of the observed population. This judgement was derived not only from the characteristics of the outliers but also from the objective of the data analysis. As was previously mentioned, it is in the nature of several (utilisation/benefit) variables to represent values that have their origins in irregular and/or catastrophic events. These sometimes quite extreme values can indeed be seen as a result of unique events, making these observations prone to be classified as outliers.

For this research these cases are a perfectly valid segment of the population and carry important and valuable information. This is because in the context of the analysis - with its objective of identifying differences between risk profiles of various population groups - a wide variation of variable values in the data was assumed. In some cases this could appear as outlier behaviour. Finally, the deletion of outliers may improve the fit of the multivariate analysis but at the same time limit its generalisability (see also Hair et al. 1998). On the basis of these arguments it could be not justified to exclude any other observations with potential outlier characteristics from the analysis, aside from the 'false negative' values.

Missing values are very common in longitudinal data because, typically, repeated measurements for one subject have to be obtained. The circumstances for obtaining these measurements are often not within the researcher's control so the results are unbalanced or partially incomplete data. The response from a subject, for example, may be missing at one point in time due to factors unrelated to the outcome of interest. Missing data affect

[184] Another diagnostic method identifies multivariate outliers. This multivariate assessment of each observation across a set of variables measures the multidimensional position of each variable relative to some common point. The Mahalanobis distance or Wilks' test are the most widely used methods for this outlier detection. Mahalanobis D2 evaluates the distance in multidimensional space of each observation from the mean center of the observation and also allows a statistical significance testing (Hair et al. 1998). The power of these classical methods decreases radically with the increasing number of observations and often fails because they are affected by the observations they are supposed to identify. Hadi (1992, 1994) developed a method that presents a reasonable solution to these problems (for the STATA™ command that executes this test see STATA 1999). However, this approach is computationally extremely demanding and when applied to the data exceeded the programs capacities. When applied to a subsample of data this diagnostic test classified a number of observations as potential outliers. However, each variable in the data shows some extreme observations, but observations seem not to have extreme patterns for an extensive number of variables. This fact alone already suggests that they cannot be considered as unrepresentative of the population. Thus even with the multivariate test separating a number of observations as outliers they seem to be similar enough to the rest of the observations. It was decided that no observation should be eliminated on the grounds of this diagnostic outlier test.

the generalisability of the results and the primary concern is to determine the reasons or mechanisms underlying the missing data. The extent of missing data is mostly a secondary issue (Hair et al. 1998).

All missing data processes lead to missing values, either due to events external to the respondent (like data collection problems and entry errors) or due to the respondent's behaviour (like information denial). If the missing data process can be explicitly identified the missing data are called ignorable, because the use of specialised analysis methods will accommodate these data. Ignorable missing data processes include, for instance, missing observations from a population that are not included in a taken sample or censored data.

Censored data are incomplete because of their stage in the missing data process (Hair et al. 1998). In the thesis this type of ignorable missing data occurs. To recall, households were divided into groups. The control group consists of households that stayed insured throughout the whole four-year observation period. On the other hand the mover groups of leavers, joiners, and switchers consisted of households that naturally could have joined, left, or switched during any year within the four-year observation period. For these movers the data entries for one, two, and even three years can be missing, because they left, switched, or did not yet join. However, the techniques and software procedures selected for this thesis' censored data (through generalised estimating equation method, GEE, for generalised linear models, GLM) are robust to this problem. Thus while ignorable missing data are a result of random missing data processes, where the observed values are a random sample of population values, they can be explicitly accommodated by selected analytical methods.

However, most missing data processes cannot be addressed with particular analytical techniques. Therefore, it is first necessary to assess the extent and impact of the missing data in order to determine whether or not the missing data process is random and, if not, which approaches for dealing with missing data should be applied (Hair et al. 1998).

The following table A3-2 gives the summary statistics for all missing values in the original data. The second column describes the number and proportion of cases with missing values for each variable. The last set of columns displays the number and percent of valid cases including mean and standard variation for all metric variables. Missing values range from zero to 10 percent of all cases for a given variable. The optical benefits variable has the highest level of missing values of around ten percent. Most other variables lie well below that, at around five percent. Among the N=1,011,735 observations

of the data 833,755 have not a single missing value for any variable. This means 88% of all observations in the data are complete.

Table A3-2: Summary statistics and pattern of missing data

Variable	Missing values		Valid cases			
	Number	Percent	Number	Percent	Mean	Std.dev.
age	2,346	0.25	945,155	99.75	41.57	13.06
gender (female)	0	0	947,501	100	n.a.	n.a.
dependants	0	0	947,501	100	1.90	1.68
income	0	0	947,501	100	551.61	341.86
race	0	0	947,501	100	n.a.	n.a.
tac (claims)	40	0	947,461	99.98	248.98	588.92
tco (contributions)	0	0	947,501	100	242.63	122.56
co2 (co-payments)	0	0	947,501	99.80	25.30	145.95
gp (GP benefits)	52,974	5.59	894,527	94.41	23.43	21.40
pub (pub.-hospital benefits)	65,657	6.93	881,844	93.07	2.27	50.42
spe (specialist benefits)	52,995	5.59	894,506	94.41	42.79	127.55
prv (priv. Hospital benefits)	56,853	6.00	890,648	94.00	55.36	325.03
den (dental benefits)	53,170	5.61	894,331	94.39	17.53	37.98
opt (optical benefits)	93,771	9.90	853,730	90.10	10.70	21.50
par (paramedical benefits)	80,988	8.55	866,513	91.45	3.64	32.65

Note: This is based on the original data of N=1,011,735 observations over n=353,458 households.

It is most likely that data entry mistakes are the reason for the manifestation of missing values in this data. For example, in the case of relatively rare medical benefits, like paramedical or optical benefits, when the actual benefit was zero the entry was often falsely left empty or filled with a dot in the spreadsheet and not filled with a numerical zero. When transforming these data from spreadsheets into STATA™, STATA™ automatically fills empty cells with a dot and counts all cells with a dot as missing data values. In summary, the missing data process appears to be due to procedural factors such as data error entries that create invalid codes, disclosure restrictions, failure to complete entries or the morbidity of a respondent. This type of missing data process cannot be explicitly identified. It is thus not ignorable. However, the classification of the missing data process in regard to its degree of randomness would allow the application of appropriate approaches to deal with missing data.

The randomness classification of missing data processes described by Little and Rubin (1987) is most widely accepted. It describes three missing data mechanisms, which consider the response variable and the effect of the covariate. First, there is the missing completely at random (MCAR) mechanism if the probability of observing the response is independent of observed and unobserved outcome values and the covariate. Second, there

is the missing at random (MAR) mechanism if the probability of observing the response is independent of the unobserved outcome values but dependent on the observed outcome values and the covariate, conditioned on the value of the covariate. Third, the missing data mechanism is non-random or informative if the probability of observing the response depends on unobserved outcome values and possibly the covariate (see Davis 2002 and Diggle et al. 2002).[185]

There are different methods for the diagnosis of the randomness of the missing data processes. The most widely used approach assesses the randomness of missing data processes through group comparisons of observations with missing versus valid data for each variable on the other variables. This standard approach is used here in order to classify the missing data process.[186] For this approach each variable's observations with valid values in one group and observations with missing values in the other group are tested for significant differences on the remaining metric variables. The process starts with the test for group differences of the first variable on the remaining variables. Subsequently the process is iterated for every variable (i.e. for each variable these groups are formed and then tested for significant differences on the remaining variables), until it is completed for all variables. The missing data process can be assumed to be MCAR if no significant differences between the two groups on other variables exist. For metric variables *t*-tests can be performed. For categorical variables a comparison of the proportions of each group is done. The results of this standard test for missing data randomness are shown in the tables A3-3 to A3-4 at the end of this section.

From these it could be concluded that the missing data process for this thesis' data is MCAR or at least MAR. Among the twenty-six variables twenty were metric and six categorical. Three metric variables - dependants, income and contributions - had no missing values and thus could not be tested against the other variables. Of the remaining 17 metric and six categorical variables, the majority showed a clear random missing data process with mostly insignificant differences between the groups of missing and non-missing values of the other variables. A noticeable pattern of predominantly significant

[185] Diggle and Kenward (1994) refer to the MCAR as complete random dropouts, the MAR as random dropouts and the non-random mechanism as the informative dropout, but consider only the response variable.
[186] A second less often used approach for assessing the randomness of missing data is through dichotomised variable correlation. The dichotomous variables are formed by replacing valid values with one values and missing values with zero values. Then the correlation between missing values of pairs of variables is analysed. Missing data processes are not associated, i.e. do not affect a number of variables and thus are MCAR if the correlation is low or moderate. This test works, however, only for metric variables. A third method, Little's MCAR test is an overall test of randomness for missing data that determines whether the

tests between the two groups only occurred for one variable - age. However, since the number of cases with missing values for this variable is extremely low (on average around 5% of all recorded values) the impact of these differences is likely to be marginal. The small number of cases involved makes these significant *t*-tests of little concern in regard to a non-random missing data process. The observed values of the variables can be seen as a true random sample of all values with no underlying process that biases the observed data. In summary, with 88% of all cases complete (without any missing values), and the majority of missing data concentrated in a very small proportion of cases, the missing data process is assumed to be random and is classified as MCAR or at least MAR. This classification of the missing data process allows the application of a range of approaches to deal with missing data.

Missing data mechanisms that are classified as MCAR allow the application of standard approaches to analysis and the use of one of the following remedies for the missing data. These are: (1) use of observations with complete data only, (2) deletion of cases and/or variables, (3) imputation, and (4) model based procedures (Hair et al 1998, Little and Rubin 1987).[187] In the first, complete case approach, all cases with missing data are discarded. In the second approach certain subsets of cases or variables that have unusually high concentrations of missing values are deleted. Imputation means that the missing values are estimated based on valid values of other variables and/or cases in the sample.

The literature is divided on the preference of one or the other approach of dealing with missing data. In case of randomly missing data Hisao (1998) and Ramsay (2002) recommend the complete case approach in favour of any other method. Ramsay (2002) argues that the best option is to live with bias problems of complete data subsamples rather than accepting the problems that data deletion and imputation bring. The main argument against the complete case approach is that it is wasteful of data when the dropout process is unrelated to the measurement process[188] and in many situations the sample size might be reduced to an inappropriate size. Hair et al. (1998) and Diggle et al.

actual missing data pattern is significantly different from expected randomly distributed missing data. Unfortunately the test is not part of the STATA™ software and could not be applied here.

[187] If the missing data mechanism is MAR the non-response mechanism is ignorable and likelihood based inferences are still valid. In the case of the non-ignorable missing data mechanism all methods of analysis are biased and even though some approaches for these cases have been developed most of them are not yet available in standard software packages (Davis 2002).

(2002) prefer imputation methods as the solution for data with MCAR processes. However, careful consideration has to be given to its potential impact on the analysis, in particular to the seduction of believing that the data are complete after all and to the danger that estimators applied for real and imputed data are substantially biased (Dempster and Rubin 1983).

The simplest solutions for this thesis' missing data process, classified as MCAR or at least MAR, are either the complete case analysis or the deletion of cases and/or variables. The descriptive statistics for the missing values in the South African data, as discussed above, showed that the extent of missing data is small and that the sample is sufficiently large to allow the deletion of all cases with missing data. For the data analysis of this thesis the complete case analysis was chosen, i.e. all cases with missing data were excluded.

[188] If the data was not MCAR, it would possess non-random elements that bias results and even though only complete observations are used, it cannot be assumed a random sample with respect to the distribution of the measurements of y_{it}. The results would be not generalisable to the population.

Table A3-3: Results for missing data randomness diagnosis through group comparisons with missing versus valid data for categorical variables

Groups formed by missing/valid data

Variable Female

	age				*tac*				*den*				*gp*				*spe*		
genF	0	1	Total	genF	0	1	Total	genF	0	1	Total	genF	0	1	Total	genF	0	1	Total
0	1,904	704,984	706,888	0	158	706,730	706,888	0	34,450	672,438	706,888	0	34,260	672,628	706,888	0	34,392	672,496	706,888
	0.19	69.81	70		0.02	69.99	70		3.41	66.59	70		3.39	66.61	70		3.41	66.6	70
1	560	302,356	302,916	1	61	302,855	302,916	1	18,754	284,162	302,916	1	18,724	284,192	302,916	1	18,757	284,159	302,916
	0.06	29.94	30		0.01	29.99	30		1.86	28.14	30		1.85	28.14	30		1.86	28.14	30
Total	2,464	1,007,340	1,009,804	Total	219	1,009,585	1,009,804	Total	53,204	956,600	1,009,804	Total	52,984	956,820	1,009,804	Total	53,149	956,655	1,009,804
	0.24	99.76	100		0.02	99.98	100		5.27	94.73	100		5.25	94.75	100		5.26	94.74	100
	opt				*par*				*co2*				*prv*				*pub*		
genF	0	1	Total	genF	0	1	Total	genF	0	1	Total	genF	0	1	Total	genF	0	1	Total
0	69,561	637,327	706,888	0	58,721	648,167	706,888	0	1,686	705,202	706,888	0	37,868	669,020	706,888	0	45,540	661,348	706,888
	6.89	63.11	70		5.82	64.19	70		0.17	69.84	70		3.75	66.25	70		4.51	65.49	70
1	27,349	275,567	302,916	1	24,462	278,454	302,916	1	474	302,442	302,916	1	19,451	283,465	302,916	1	21,163	281,753	302,916
	2.71	27.29	30		2.42	27.58	30		0.05	29.95	30		1.93	28.07	30		2.1	27.9	30
Total	96,910	912,894	1,009,804	Total	83,183	926,621	1,009,804	Total	2,160	1,007,644	1,009,804	Total	57,319	952,485	1,009,804	Total	66,703	943,101	1,009,804
	9.6	90.4	100		8.24	91.76	100		0.21	99.79	100		5.68	94.32	100		6.61	93.39	100

Urban

	age				*tac*				*den*				*gp*				*spe*		
urban	0	1	Total	urban	0	1	Total	urban	0	1	Total	urban	0	1	Total	urban	0	1	Total
0	123	67,320	67,443	0	37	67,406	67,443	0	4,567	62,876	67,443	0	4,559	62,884	67,443	0	4,566	62,877	67,443
	0.01	6.67	6.68		0	6.68	6.68		0.45	6.23	6.68		0.45	6.23	6.68		0.45	6.23	6.68
1	2,341	940,020	942,361	1	182	942,179	942,361	1	48,637	893,724	942,361	1	48,425	893,936	942,361	1	48,583	893,778	942,361
	0.23	93.09	93.32		0.02	93.3	93.32		4.82	88.5	93.32		4.8	88.53	93.32		4.81	88.51	93.32
Total	2,464	1,007,340	1,009,804	Total	219	1,009,585	1,009,804	Total	53,204	956,600	1,009,804	Total	52,984	956,820	1,009,804	Total	53,149	956,655	1,009,804
	0.24	99.76	100		0.02	99.98	100		5.27	94.73	100		5.25	94.75	100		5.26	94.74	100
	opt				*par*				*co2*				*prv*				*pub*		
urban	0	1	Total	urban	0	1	Total	urban	0	1	Total	urban	0	1	Total	urban	0	1	Total
0	6,552	60,891	67,443	0	5,902	61,541	67,443	0	91	67,352	67,443	0	4,749	62,694	67,443	0	5,176	62,267	67,443
	0.65	6.03	6.68		0.58	6.09	6.68		0.01	6.67	6.68		0.47	6.21	6.68		0.51	6.17	6.68
1	90,358	852,003	942,361	1	77,281	865,080	942,361	1	2,069	940,292	942,361	1	52,570	889,791	942,361	1	61,527	880,834	942,361
	8.95	84.37	93.32		7.65	85.67	93.32		0.2	93.12	93.32		5.21	88.12	93.32		6.09	87.23	93.32
Total	96,910	912,894	1,009,804	Total	83,183	926,621	1,009,804	Total	2,160	1,007,644	1,009,804	Total	57,319	952,485	1,009,804	Total	66,703	943,101	1,009,804
	9.6	90.4	100		8.24	91.76	100		0.21	99.79	100		5.68	94.32	100		6.61	93.39	100

Cont.

Cont.

Asian

ra	age 0	age 1	age Total	ra	tac 0	tac 1	tac Total	ra	den 0	den 1	den Total	ra	gp 0	gp 1	gp Total	ra	spe 0	spe 1	spe Total
0	2,416	919,019	921,435	0	185	921,250	921,435	0	51,035	870,400	921,435	0	50,826	870,609	921,435	0	50,982	870,453	921,435
	0.24	91.01	91.25		0.02	91.23	91.25		5.05	86.19	91.25		5.03	86.22	91.25		5.05	86.2	91.25
1	48	88,321	88,369	1	34	88,335	88,369	1	2,169	86,200	88,369	1	2,158	86,211	88,369	1	2,167	86,202	88,369
	0	8.75	8.75		0	8.75	8.75		0.21	8.54	8.75		0.21	8.54	8.75		0.21	8.54	8.75
Total	2,464	1,007,340	1,009,804	Total	219	1,009,585	1,009,804	Total	53,204	956,600	1,009,804	Total	52,984	956,820	1,009,804	Total	53,149	956,655	1,009,804
	0.24	99.76	100		0.02	99.98	100		5.27	94.73	100		5.25	94.75	100		5.26	94.74	100

ra	opt 0	opt 1	opt Total	ra	par 0	par 1	par Total	ra	co2 0	co2 1	co2 Total	ra	prv 0	prv 1	prv Total	ra	pub 0	pub 1	pub Total
0	90,425	831,010	921,435	0	78,116	843,319	921,435	0	1,992	919,443	921,435	0	54,824	866,611	921,435	0	63,364	858,071	921,435
	8.95	82.29	91.25		7.74	83.51	91.25		0.2	91.05	91.25		5.43	85.82	91.25		6.27	84.97	91.25
1	6,485	81,884	88,369	1	5,067	83,302	88,369	1	168	88,201	88,369	1	2,495	85,874	88,369	1	3,339	85,030	88,369
	0.64	8.11	8.75		0.5	8.25	8.75		0.02	8.73	8.75		0.25	8.5	8.75		0.33	8.42	8.75
Total	96,910	912,894	1,009,804	Total	83,183	926,621	1,009,804	Total	2,160	1,007,644	1,009,804	Total	57,319	952,485	1,009,804	Total	66,703	943,101	1,009,804
	9.6	90.4	100		8.24	91.76	100		0.21	99.79	100		5.68	94.32	100		6.61	93.39	100

Black

rb	age 0	age 1	age Total	rb	tac 0	tac 1	tac Total	rb	den 0	den 1	den Total	rb	mgp 0	mgp 1	mgp Total	rb	spe 0	spe 1	spe Total
0	1,947	728,683	730,630	0	191	730,439	730,630	0	35,123	695,507	730,630	0	34,925	695,705	730,630	0	35,076	695,554	730,630
	0.19	72.16	72.35		0.02	72.33	72.35		3.48	68.88	72.35		3.46	68.9	72.35		3.47	68.88	72.35
1	517	278,657	279,174	1	28	279,146	279,174	1	18,081	261,093	279,174	1	18,059	261,115	279,174	1	18,073	261,101	279,174
	0.05	27.6	27.65		0	27.64	27.65		1.79	25.86	27.65		1.79	25.86	27.65		1.79	25.86	27.65
Total	2,464	1,007,340	1,009,804	Total	219	1,009,585	1,009,804	Total	53,204	956,600	1,009,804	Total	52,984	956,820	1,009,804	Total	53,149	956,655	1,009,804
	0.24	99.76	100		0.02	99.98	100		5.27	94.73	100		5.25	94.75	100		5.26	94.74	100

rb	opt 0	opt 1	opt Total	rb	par 0	par 1	par Total	rb	co2 0	co2 1	co2 Total	rb	prv 0	prv 1	prv Total	rb	pub 0	pub 1	pub Total
0	73,785	656,845	730,630	0	61,717	668,913	730,630	0	1,552	729,078	730,630	0	38,807	691,823	730,630	0	47,201	683,429	730,630
	7.31	65.05	72.35		6.11	66.24	72.35		0.15	72.2	72.35		3.84	68.51	72.35		4.67	67.68	72.35
1	23,125	256,049	279,174	1	21,466	257,708	279,174	1	608	278,566	279,174	1	18,512	260,662	279,174	1	19,502	259,672	279,174
	2.29	25.36	27.65		2.13	25.52	27.65		0.06	27.59	27.65		1.83	25.81	27.65		1.93	25.72	27.65
Total	96,910	912,894	1,009,804	Total	83,183	926,621	1,009,804	Total	2,160	1,007,644	1,009,804	Total	57,319	952,485	1,009,804	Total	66,703	943,101	1,009,804
	9.6	90.4	100		8.24	91.76	100		0.21	99.79	100		5.68	94.32	100		6.61	93.39	100

Cont.

Cont. **Coloured**

	age				tac				den				mgp				spe		
rc	0	1	Total	rc	0	1	Total	rc	0	1	Total	rc	0	1	Total	rc	0	1	Total
0	1,752	677,852	679,604	0	144	679,460	679,604	0	39,521	640,083	679,604	0	39,340	640,264	679,604	0	39,409	640,195	679,604
	0.17	67.13	67.3		0.01	67.29	67.3		3.91	63.39	67.3		3.9	63.4	67.3		3.9	63.4	67.3
1	712	329,488	330,200	1	75	330,125	330,200	1	13,683	316,517	330,200	1	13,644	316,556	330,200	1	13,740	316,460	330,200
	0.07	32.63	32.7		0.01	32.69	32.7		1.36	31.34	32.7		1.35	31.35	32.7		1.36	31.34	32.7
Total	2,464	1,007,340	1,009,804	Total	219	1,009,585	1,009,804	Total	53,204	956,600	1,009,804	Total	52,984	956,820	1,009,804	Total	53,149	956,655	1,009,804
	0.24	99.76	100		0.02	99.98	100		5.27	94.73	100		5.25	94.75	100		5.26	94.74	100
	opt				par				co2				prv				pub		
rc	0	1	Total	rc	0	1	Total	rc	0	1	Total	rc	0	1	Total	rc	0	1	Total
0	69,607	609,997	679,604	0	60,245	619,359	679,604	0	1,684	677,920	679,604	0	42,584	637,020	679,604	0	48,983	630,621	679,604
	6.89	60.41	67.3		5.97	61.33	67.3		0.17	67.13	67.3		4.22	63.08	67.3		4.85	62.45	67.3
1	27,303	302,897	330,200	1	22,938	307,262	330,200	1	476	329,724	330,200	1	14,735	315,465	330,200	1	17,720	312,480	330,200
	2.7	30	32.7		2.27	30.43	32.7		0.05	32.65	32.7		1.46	31.24	32.7		1.75	30.94	32.7
Total	96,910	912,894	1,009,804	Total	83,183	926,621	1,009,804	Total	2,160	1,007,644	1,009,804	Total	57,319	952,485	1,009,804	Total	66,703	943,101	1,009,804
	9.6	90.4	100		8.24	91.76	100		0.21	99.79	100		5.68	94.32	100		6.61	93.39	100

White

	age				tac				den				gp				spe		
rw	0	1	Total	rw	0	1	Total	rw	0	1	Total	rw	0	1	Total	rw	0	1	Total
0	1,277	696,466	697,743	0	137	697,606	697,743	0	33,933	663,810	697,743	0	33,861	663,882	697,743	0	33,980	663,763	697,743
	0.13	68.97	69.1		0.01	69.08	69.1		3.36	65.74	69.1		3.35	65.74	69.1		3.37	65.73	69.1
1	1,187	310,874	312,061	1	82	311,979	312,061	1	19,271	292,790	312,061	1	19,123	292,938	312,061	1	19,169	292,892	312,061
	0.12	30.79	30.9		0.01	30.9	30.9		1.91	28.99	30.9		1.89	29.01	30.9		1.9	29	30.9
Total	2,464	1,007,340	1,009,804	Total	219	1,009,585	1,009,804	Total	53,204	956,600	1,009,804	Total	52,984	956,820	1,009,804	Total	53,149	956,655	1,009,804
	0.24	99.76	100		0.02	99.98	100		5.27	94.73	100		5.25	94.75	100		5.26	94.74	100
	opt				par				co2				prv				pub		
rw	0	1	Total	rw	0	1	Total	rw	0	1	Total	rw	0	1	Total	rw	0	1	Total
0	56,913	640,830	697,743	0	49,471	648,272	697,743	0	1,252	696,491	697,743	0	35,742	662,001	697,743	0	40,561	657,182	697,743
	5.64	63.46	69.1		4.9	64.2	69.1		0.12	68.97	69.1		3.54	65.56	69.1		4.02	65.08	69.1
1	39,997	272,064	312,061	1	33,712	278,349	312,061	1	908	311,153	312,061	1	21,577	290,484	312,061	1	26,142	285,919	312,061
	3.96	26.94	30.9		3.34	27.56	30.9		0.09	30.81	30.9		2.14	28.77	30.9		2.59	28.31	30.9
Total	96,910	912,894	1,009,804	Total	83,183	926,621	1,009,804	Total	2,160	1,007,644	1,009,804	Total	57,319	952,485	1,009,804	Total	66,703	943,101	1,009,804
	9.6	90.4	100		8.24	91.76	100		0.21	99.79	100		5.68	94.32	100		6.61	93.39	100

Note: Each cell contains the table for the categorical variable of each page: (1) female gender, (2) urban residency, (3) Asian, (4) Black, (5) Coloured or (6) White race and the comparison of its realisation frequency within the group of missing and valid values of the other variables. The first row for each variable realisation in the cells contains the number of observations of the row variable for the column variable groups of missing and valid data. The rows below display the proportion. The interpretation of the bottom left cell is the following: the proportion of missing values for the variable co-payment2 among Whites is 0.09% in comparison to the proportion among Non-Whites 0.12%. This result leads to the conclusion that there is no significant difference between these proportions. Moreover, the very few cases of missing values for co-payment2 of 908 among Whites and 1,252 among Non-Whites make it extremely unlikely that their ignorable difference has any impact in form of a non-random missing data process.

Table A3-4: T-test results for missing data randomness diagnosis through group comparisons with missing versus valid data for metric variables

Variables	Groups formed by missing/valid data	age	den	gp	opt	par	prv	pub	spe	tac	co2
age	Sig-level	\	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
nrandom	Missing	\	53,107	52,887	96,698	83,006	57,216	66,570	53,051	218	2,159
	Present	\	954,233	954,453	910,642	924,334	950,124	940,770	954,289	1,007,122	1,005,181
	m missing	\	37.52	37.49	40.58	39.95	38.06	38.93	37.50	46.97	45.18
	m present	\	41.73	41.73	41.61	41.65	41.71	41.69	41.73	41.51	41.50
den	Sig-level	0.094	\	0.676	0.000	0.000	0.000	0.000	0.879	0.698	0.701
random	missing	2,367	\	9	43,722	29,996	4,135	13,518	148	177	2,159
	present	954,233	\	956,591	912,878	926,604	952,465	943,082	956,452	956,423	954,441
	m missing	18.87	\	22.89	32.50	34.10	38.16	36.35	17.07	16.44	17.87
	m present	17.55	\	17.56	16.84	17.02	17.47	17.29	17.56	17.56	17.55
gp	Sig-level	0.641	0.000	\	0.000	0.000	0.000	0.000	0.079	0.075	0.183
r/nrandom	missing	2,367	229	\	43,935	30,209	4,345	13,729	175	177	2,160
	present	954,453	956,591	\	912,885	926,611	952,475	943,091	956,645	956,643	954,660
	m missing	23.73	41.43	\	32.65	33.33	35.46	34.28	26.42	20.61	22.90
	m present	23.52	23.52	\	23.08	23.20	23.47	23.37	23.52	23.52	23.52
opt	Sig-level	0.010	0.416	0.347	\	0.464	0.877	0.646	0.470	0.218	0.000
random	missing	2,252	16	9	\	34	132	64	129	156	1,884
	present	910,642	912,878	912,885	\	912,860	912,762	912,830	912,765	912,738	911,010
	m missing	11.84	15.06	17.45	\	7.96	10.96	9.43	12.05	8.54	7.92
	m present	10.67	10.67	10.67	\	10.67	10.67	10.67	10.67	10.67	10.68
par	Sig-level	0.000	0.644	0.721	0.000	\	0.114	0.509	0.883	0.770	0.000
random	missing	2,287	17	10	13,761	\	147	64	134	163	1,945
	present	924,334	926,604	926,611	912,860	\	926,474	926,557	926,487	926,458	924,676
	m missing	8.20	0.03	0.00	17.48	\	8.19	1.03	4.23	4.58	9.46
	m present	3.80	3.81	3.81	3.60	\	3.81	3.81	3.81	3.81	3.80
prv	Sig-level	0.002	0.795	0.052	0.000	0.000	\	0.000	0.537	0.219	0.000
r/nrandom	missing	2,361	20	10	39,723	26,011	\	9,544	144	112	2,127
	present	950,124	952,465	952,475	912,762	926,474	\	942,941	952,341	952,373	950,358
	m missing	79.74	38.42	263.91	302.63	291.86	\	249.40	75.10	19.00	168.83
	m present	57.81	57.86	57.86	47.21	51.29	\	55.93	57.86	57.87	57.62

Cont.

Cont.

pub	Sig-level	0.028	0.906	0.881	0.000	0.000	0.926	\	0.722	0.473	0.000
random	missing	2,331	19	10	30,271	16,544	160	\	141	167	2,060
	present	940,770	943,082	943,091	912,830	926,557	942,941	\	942,960	942,934	941,041
	m missing	4.79	3.84	0.00	8.85	10.32	2.07	\	3.99	5.31	8.79
	m present	2.44	2.44	2.44	2.23	2.30	2.44	\	2.44	2.44	2.43
spe	Sig-level	0.000	0.000	0.050	0.000	0.000	0.000	0.000	\	0.082	0.000
r/nrandom	missing	2366	203	10	43890	30168	4314	13695	\	135	2156
	present	954289	956452	956645	912765	926487	952341	942960	\	956520	954499
	m missing	56.27	317.19	161.68	169.13	175.98	225.17	199.18	\	63.61	122.43
	m present	43.79	43.76	43.82	37.79	39.52	43.00	41.56	\	43.82	43.64
tac	Sig-level	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	\	0.000
r/nrandom	missing	2,463	53,162	52,942	96,847	83,127	57,212	66,651	53,065	\	2,093
	present	1,007,122	956,423	956,643	912,738	926,458	952,373	942,934	956,520	\	1,007,492
	m missing	327.65	10.02	0.33	398.21	332.12	92.12	213.68	2.06	\	480.42
	m present	252.12	265.78	266.25	236.83	245.15	261.93	255.04	266.19	\	251.83
co2	Sig-level	0.252	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.739	\
r/nrandom	missing	2,463	53,203	52,984	96,634	82,968	57,286	66,603	53,145	152	\
	present	1,005,181	954,441	954,660	911,010	924,676	950,358	941,041	954,499	1,007,492	\
	m missing	30.58	1.18	0.02	42.89	36.84	10.49	24.28	0.30	22.69	\
	m present	26.94	28.38	28.44	25.25	26.06	27.94	27.13	28.43	26.95	\

Note: Each cell contains five values. The first is the significance for the t-test for comparing the row variable mean for the column variable groups of missing data versus valid data. The second and third cells contain the number of observations of the row variable for the column variable groups of missing and valid data. The last two cells display the mean of the row variable for the column variable groups of missing and valid data. The interpretation of the bottom left cell is the following: the t-test comparing the mean for the row variable co-payment2 between the groups of missing and valid data of the column variable age has a significance level of 0.252. This result leads to intermediary conclusion that there is no significant difference between the mean of co-payment2 for the group of missing and valid values in age. Moreover, the sample sizes of the co-payment2 variable for the groups of missing values and valid values in age are 2,463 and 1,005,181, respectively. The extremely small sample size for the missing values (0.24% of all recorded values) indicates that the impact of any differences would be marginal and would give no reason for concern, even if the t-test would have been significant. The mean of co-payment2 for the group of missing values in age is 30.6 Rand per person per month, for the group of valid values 26,9.

Appendix 4

Medical schemes and their plan options

Table A4-1: Medical schemes by number of observations in ascending order, 1995-1998

Rank	Total frequency over the four-year observation period *f(N)*	Average frequency over the four-year observation period *f(N/4)*	Percent	Cum. (%)
29	2,427	607	0.24	0.24
28	2,987	747	0.30	0.54
27	3,153	788	0.31	0.85
26	6,831	1,708	0.68	1.52
25	6,876	1,719	0.68	2.20
24	7,650	1,913	0.76	2.96
23	7,709	1,927	0.76	3.72
22	7,809	1,952	0.77	4.49
21	8,147	2,037	0.81	5.30
20	9,210	2,303	0.91	6.21
19	9,497	2,374	0.94	7.15
18	10,223	2,556	0.94	8.08
17	11,150	2,788	0.94	9.02
16	11,154	2,789	1.10	10.13
15	12,035	3,009	1.19	11.32
14	13,613	3,403	1.35	12.66
13	15,112	3,778	1.49	14.15
12	18,202	4,551	1.80	15.95
11	18,614	4,654	1.84	17.79
10	21,333	5,333	2.11	19.90
9	21,539	5,385	2.13	22.03
8	35,660	8,915	3.52	25.56
7	38,463	9,616	3.80	29.36
6	42,043	10,511	4.16	33.51
5	61,805	15,451	6.11	39.62
4	68,155	17,039	6.74	46.36
3	80,582	20,146	7.96	54.32
2	180,982	45,246	17.89	72.21
1	278,774	69,694	27.55	100
Total	**1,011,735**	**252,934**	**100**	

Note: Closed schemes are marked black, open schemes red. The numbers are based on the original data over N=1,011,735 observations and n=353,458 households. For reasons of confidentiality the names of the schemes are not given here. However, upon request they are available from the author.

Table A4-2: Medical schemes and their plan options by frequency in ascending order, 1995-1998

Rank of scheme	Number of plans	Average frequency *f(N/1995-1998)*	Percent	Cum.
29	1	607	0.24	0.24
28	2	747	0.30	0.54
27	3	788	0.31	0.85
26	4	1,708	0.68	1.52
25	5	1,719	0.68	2.20
24	6	1,913	0.76	2.96
23	7	1,927	0.76	3.72
22	8	1,952	0.77	4.49
21	9	2,037	0.81	5.30
20	10	2,303	0.91	6.21
19	11	2,374	0.94	7.15
18	12	2,556	1.01	8.16
17	13	2,788	1.10	9.26
16		2,789	1.10	10.36
	14	1,418	0.56	9.82
	15	632	0.25	10.07
	16	739	0.29	10.36
15		3,009	1.19	11.55
	17	2,749	1.09	11.45
	18	260	0.10	11.55
14	19	3,403	1.35	12.90
13	20	3,778	1.49	14.39
12		4,551	1.80	16.19
	21	1,097	0.43	14.82
	22	3,453	1.37	16.19
11	23	4,654	1.84	18.03
10	24	5,333	2.11	20.14
9	25	5,385	2.13	22.27
8		8,915	3.52	25.79
	26	4,412	1.74	24.01
	27	1,630	0.64	24.65
	28	1,300	0.51	25.17
	29	1,574	0.62	25.79
7		9,616	3.80	29.59
	30	351	0.14	25.93
	31	1,564	0.62	26.55
	32	2,244	0.89	27.43
	33	5,457	2.16	29.59
6		10,511	4.16	33.75
	34	6,416	2.54	32.13
	35	4,095	1.62	33.75
5		15,451	6.11	39.86
	36	14,603	5.77	39.52
	37	849	0.34	39.86
4	38	17,039	6.74	46.59
3		20,146	7.96	54.56
	39	5,408	2.14	48.73
	40	3,612	1.43	50.16
	41	11,126	4.40	54.56
2		45,246	17.89	72.45
	42	4,912	1.94	56.50
	43	34,169	13.51	70.01
	44	1,067	0.42	70.43
	45	2,682	1.06	71.49
	46	1,324	0.52	72.01
	47	1,091	0.43	72.45
1		69,694	27.55	100
	48	68,709	27.16	99.61
	49	985	0.39	100
29	Total: 49	252,934	100	

Note: Closed schemes are marked black, open schemes red. The numbers are based on the original data over N=1,011,735 observations and n=353,458 households over four years of observation. The scheme and plan option names are here not disclosed for reasons of confidentiality. However, upon request they are available from the author. The table can be read like the following: The largest scheme with rank one was an open scheme and accounted on average for 69,694 (or 28%) of all observations per year. This scheme had two plan options (48 and 49), one with an annual average of 68,709 (or 27%) of all observations and one with on average 985 (or 0.4%) of all observations.

Appendix 5

Plan classification

Table A5-1: Mean intervals for of plan classification and sensitivity analysis

Variable	Mean	Classification mean interval	Sensitivity analysis 10% increased mean interval	Mean class. changes with 10% increase	Sensitivity analysis 10% decreased mean interval	Mean class. changes with 10% decrease
Socio-demographic						
age	42	38<42<46	37.6<42<46.4	1	38.4<42<45.6	3
Claims, contribution and benefits in Rand						
contributions	242	142<242<342	132<242<352	3	152<242<332	3
co-payment2	23	13<23<33	12<23<34	0	14<23<32	1
claims	229	149<229<309	141<229<317	2	157<229<301	4
Basic benefits						
GP	23	18<23<28	17.5<23<28.5	4	18.5<23<27.5	1
Public hospital	2.0	0.5<2.0<2.5	0.35<2.0<2.65	2	0.65<2.0<2.35	7
Negotiable benefits						
Specialist	36	21<36<51	19.5<36<52.5	0	22.5<36<49.5	3
Private hospital	43	23<43<63	21<43<65	1	25<43<61	2
Discretionary benefits						
Dental	17	12<17<22	11.5<17<22.5	0	12.5<17<21.5	1
Optical	11	7<11<15	6.6<11<15.4	0	7.4<11<14.6	5
Paramedical	3.3	1.3<3.3<5.3	1.1<3.3<5.5	1	1.5<3.3<5.1	2

Note: These are the means over all N=833,755 observations and n=325,957 households.

Table A5-2: Data plan classification into low- and high-risk plans, 1995-98

data 95-98	age	tco	co2	tac	gp	pub	spec	prv	dent	opt	par	risk type
49	41.6	237.3	44.8	251.4	17.7	5	31.5	31.3	19.6	10.6	3.7	average
48	68	367.6	78.4	504.3	19	3.5	74.2	116.1	18.6	9.5	34	high
47	39	256.3	41.9	245.2	17.7	1.9	33.4	50.4	15.4	13.1	2.4	average
46	44.1	443.9	85.2	388.2	13.9	0.3	60.9	51.1	27.1	11.1	9.2	high
45	53.4	297.3	72.5	305.4	5.3	0.3	49	91.1	22.8	5.4	3.3	average
44	42	255.3	45.3	296	17.8	0.3	49.9	77.7	19.9	8.3	4.7	average
43	35.7	114.9	34.4	136	17.9	0.8	14.8	15.7	8.6	3	0.3	low
42	47.4	234.3	19.5	233.4	27.1	2.1	40.1	31.2	17.3	8.6	2.3	average
41	42.9	135.4	18	107.3	18.3	5.7	13.5	14.8	5	4	0.9	low
40	45.9	233.9	15.6	261.9	27.9	1.4	41.2	59.4	16.7	11	3.4	average
39	43.9	181.4	14.2	171.1	31.2	1.6	19.8	23.1	10.9	6.5	0.8	low
38	42.1	376	25.8	285.6	15.1	0.5	55	44.1	30.2	16.6	6.4	high
37	42.8	309.5	60	285.5	14.3	0.7	44.2	56.1	20	7.1	3.8	average
36	39.7	127.8	12	144.7	26.6	1	14.4	19.7	10.9	8.6	0.7	low
35	41.9	347.5	41.5	306.2	16.2	0.6	57.9	52.1	32.9	13	4.3	high
34	46.1	293.6	42.5	270.5	17.3	0.6	39.3	52.8	17.2	8.2	4.9	average
33	37.5	319.3	37.1	269.1	18.4	0.6	57.5	59	18.1	7.3	3	average
32	41.8	289.4	59.6	281.2	14.6	0.8	49.5	68.4	20.3	5.3	1.5	average
31	39.6	235.5	36.9	215.7	17.4	0.7	29.6	39.6	14.4	8.6	2.2	low
30	59	442.3	43.3	456.3	20.3	1.7	76.6	74.4	29.3	21.3	9.8	high
29	37.5	79.2	16.2	103.1	18.9	1.2	9.6	12.2	8.8	4.1	0.3	low
28	41.1	96.4	28.2	134.1	16.3	1.7	13.7	24.9	7.9	7.1	1.1	low
27	34.9	283.9	21.8	221.2	22.3	0.4	36	50.6	18	12.7	2.7	average
26	39.4	302.9	25.9	252.9	18.7	1.3	39.3	48.6	24	12.8	4.6	average
25	51.3	322.9	48.5	327.1	17.3	6	50.8	70.1	19.6	8.4	11.1	high
24	47.2	293.3	45.7	306.6	15.4	0.9	49	71	17	10.4	4.8	average
23	38.3	206.5	40.4	250.1	16.3	0.5	41.2	45.4	17.1	13.1	3.9	average
22	41.5	283.5	39.6	300.5	19.6	2	47.1	65.1	19	8.6	4.1	average
21	39.1	146.9	14.5	180.1	29.6	3.3	21.9	32.3	13.1	8.4	2	average
20	48.1	216.1	52.7	250.6	13.8	2.9	36.8	39.2	18.6	10.6	5.2	average
19	39.4	286.8	17.3	227.5	23	0.5	35	38.6	20.7	11.9	3.2	average
18	47.2	269	50.6	286.3	14.4	1.4	46.1	54.9	17.9	10.9	6.9	average
17	37.4	350.6	28.5	306.9	21.5	0.6	55.7	54.8	29.4	24	5.9	average
16	44.3	331.9	51.7	315.7	16.5	0.9	49.4	61.5	22.3	9.6	8.7	average
15	47.8	331.3	69.3	324.3	18.1	1.6	45.6	64.8	19.2	12.1	7.4	high
14	36.8	245	23.2	207.7	21.8	1.3	33.5	40.3	19.5	9.7	2.2	average
13	44.9	414.8	45.1	356.8	20	0.5	65.6	63.6	31.3	14.1	5.6	high
12	40	241.7	16.6	220.1	24.5	1.4	30.9	38.8	20.9	10.9	3	average
11	37	104.9	20.8	114.2	20.4	1.8	9.1	11.9	7.8	4.4	0.5	low
10	38.4	147.1	11.8	133.9	15.9	1.1	20.3	28.8	6	5.8	0.9	low
9	43.9	271.7	56.3	284.4	15	0.9	40	56.3	18.3	12.6	4.7	average
8	46.8	413.8	25.5	360.6	22.9	1	59.6	64.8	28.9	15.6	10	high
7	40.9	204.7	21	194.7	20.2	1.1	33.8	40	14	7.3	3.3	average
6	45.6	336.3	31.4	322.7	25.3	0.9	52.1	66.7	23.1	13.7	5.3	high
5	36.9	201.7	18.8	188.5	19.3	1.3	32.8	37.4	18.4	7.2	2.1	average
4	41.7	261.5	14.6	260.8	29.8	1.5	44.9	40.3	21.2	10.5	3.6	average
3	45.5	304.2	35.5	296.3	18.2	1.2	48.4	62.9	22.6	16.8	5.7	average
2	40.6	194.1	10.7	176.4	27	1.4	20.2	28.6	14.1	10.2	1.3	low
1	40.6	216.7	13.3	199.5	25.6	3.6	32.8	37	11.6	10.2	2	average

Table A5-3: Classification results: low-risk plans and high-risk plans, 1995-98

High risk plans

data 95-98	age	tco	co2	tac	gp	pub	spec	prv	dent	opt	par	risk type
48	68	367.6	78.4	504.3	19	3.5	74.2	116.1	18.6	9.5	34	high
46	44.1	443.9	85.2	388.2	13.9	0.3	60.9	51.1	27.1	11.1	9.2	high
38	42.1	376	25.8	285.6	15.1	0.5	56	44.1	30.2	16.8	6.4	high
36	41.9	347.5	41.8	306.2	16.2	0.6	57.9	52.1	32.9	13	4.3	high
30	58	442.3	43.3	456.3	20.3	1.7	76.6	74.4	29.3	21.3	9.8	high
29	51.3	322.9	48.5	327.1	17.3	6	50.8	70.1	19.6	8.4	11.1	high
16	47.8	331.3	69.3	324.3	18.1	1.6	45.6	64.8	19.2	12.1	7.4	high
13	44.9	414.8	45.1	356.8	20	0.5	65.8	63.6	31.3	14.1	5.5	high
8	48.8	413.8	25.6	360.8	22.9	1	59.6	64.8	28.9	15.8	10	high
6	45.6	336.3	31.4	322.7	25.3	0.9	52.1	66.7	23.1	13.7	5.3	high
Total: 10												

Low risk plans

data 95-98	age	tco	co2	tac	gp	pub	spec	prv	dent	opt	par	risk type
43	35.7	114.9	34.4	136	17.9	0.8	14.8	15.7	8.6	3	0.3	low
41	42.9	135.4	18	107.3	18.3	5.7	13.5	14.8	5	4	0.9	low
39	43.9	181.4	14.2	171.1	31.2	1.6	19.8	23.1	10.9	6.5	0.8	low
36	39.7	127.8	12	144.7	26.6	1	14.4	19.7	10.9	8.6	0.7	low
31	39.6	235.5	36.9	215.7	17.4	0.7	29.6	39.6	14.4	8.6	2.2	low
29	37.5	79.2	16.2	103.1	18.9	1.2	9.6	12.2	8.8	4.1	0.3	low
28	41.1	96.4	28.2	134.1	16.3	1.7	13.7	24.9	7.9	7.1	1.1	low
11	37	104.9	20.8	114.2	20.4	1.8	9.1	11.9	7.8	4.4	0.8	low
10	38.4	147.1	11.8	133.9	15.9	1.1	20.3	28.8	6	5.8	0.9	low
Total 10												

Table A5-4: Data plan classification into low- and high-risk plans with 10% increased mean interval, 1995-98

data 95-98	age	tco	co2	tac	gp	pub	spec	prv	dent	opt	par	risk type
49	41.6	237.3	44.8	251.4	17.7	5	31.5	31.3	19.6	10.6	3.7	average
48	68	367.6	78.4	504.3	19	3.5	74.2	116.1	18.6	9.5	34	high
47	39	256.3	41.9	245.2	17.7	1.9	33.4	50.4	15.4	13.1	2.4	average
46	44.1	443.9	85.2	388.2	13.9	0.3	60.9	51.1	27.1	11.1	9.2	high
45	53.4	297.3	72.6	305.4	5.3	0.3	49	91.1	22.8	5.4	3.3	average
44	42	255.3	45.3	296	17.8	0.3	49.9	77.7	19.9	8.3	4.7	average
43	35.7	114.9	34.4	136	17.9	0.8	14.8	15.7	8.6	3	0.3	low
42	47.4	234.3	19.5	233.4	27.1	2.1	40.1	31.2	17.3	8.6	2.3	average
41	42.9	135.4	18	107.3	18.3	5.7	13.5	14.8	5	4	0.9	low
40	45.9	233.9	15.6	261.9	27.9	1.4	41.2	59.4	16.7	11	3.4	average
39	43.9	181.4	14.2	171.1	31.2	1.6	19.8	23.1	10.9	6.5	0.8	low
38	42.1	376	25.8	285.6	15.1	0.5	55	44.1	30.2	16.6	6.4	high
37	42.8	309.5	60	285.5	14.3	0.7	44.2	56.1	20	7.1	3.8	average
36	39.7	127.8	12	144.7	26.6	1	14.4	19.7	10.9	8.6	0.7	low
35	41.9	347.5	41.5	306.2	16.2	0.6	57.9	52.1	32.8	13	4.3	high
34	46.1	293.6	42.5	270.5	17.3	0.6	39.3	52.8	17.2	8.2	4.9	average
33	37.5	319.3	37.1	269.1	18.4	0.6	57.5	59	18.1	7.3	3	average
32	41.8	289.4	59.6	281.2	14.6	0.8	49.5	68.4	20.3	5.3	1.5	average
31	39.6	235.5	36.8	215.7	17.4	0.7	29.6	39.6	14.4	8.6	2.2	low
30	59	442.3	43.3	456.3	20.3	1.7	76.6	74.4	29.3	21.3	9.8	high
29	37.5	79.2	16.2	103.1	18.9	1.2	9.6	12.2	8.8	4.1	0.3	low
28	41.1	96.4	28.2	134.1	16.3	1.7	13.7	24.9	7.9	7.1	1.1	low
27	34.9	283.9	21.8	221.2	22.3	0.4	36	50.6	18	12.7	2.7	average
26	39.4	302.9	25.9	252.9	18.7	1.3	39.3	48.6	24	12.8	4.6	average
25	51.3	322.9	48.5	327.1	17.3	6	50.8	70.1	19.6	8.4	11.1	high
24	47.2	293.3	45.7	306.6	15.4	0.9	49	71	17	10.4	4.8	average
23	38.3	206.5	40.8	250.1	16.3	0.5	41.2	45.4	17.1	13.1	3.9	average
22	41.5	283.5	39.6	300.5	19.6	2	47.1	65.1	19	8.6	4.1	average
21	39.1	146.9	14.5	180.1	29.6	3.3	21.9	32.3	13.1	8.4	2	average
20	48.1	216.1	52.7	250.6	13.8	2.9	36.8	39.2	18.6	10.6	5.2	average
19	39.4	266.8	17.3	227.5	23	0.5	35	38.6	20.7	11.9	3.2	average
18	47.2	269	50.8	286.3	14.4	1.4	46.1	54.9	17.9	10.9	6.9	average
17	37.4	350.6	28.5	306.9	21.5	0.6	55.7	54.8	29.4	24	5.9	average
16	44.3	331.9	51.7	315.7	16.5	0.9	49.4	61.5	22.3	9.6	8.7	average
15	47.8	331.3	69.3	324.3	18.1	1.6	45.6	64.8	19.2	12.1	7.4	high
14	36.8	245	23.2	207.7	21.8	1.3	33.5	40.3	19.5	9.7	2.2	average
13	44.9	414.8	45.1	356.5	20	0.5	65.6	63.6	31.3	14.1	5.5	high
12	40	241.7	16.6	220.1	24.5	1.4	30.9	38.8	20.9	10.9	3	average
11	37	104.9	20.8	114.2	20.4	1.8	9.1	11.9	7.8	4.4	0.5	low
10	38.4	147.1	11.8	133.9	15.9	1.1	20.3	28.8	6	5.8	0.9	low
9	43.9	271.7	56.3	284.4	15	0.9	40	56.3	18.3	12.6	4.7	average
8	46.8	413.8	25.5	360.6	22.9	1	59.6	64.8	26.9	15.6	10	high
7	40.9	204.7	21	194.7	20.2	1.1	33.8	40	14	7.3	3.3	average
6	45.6	336.3	31.4	322.7	25.3	0.9	52.1	66.7	23.1	13.7	5.3	high
5	36.9	201.7	18.8	188.5	19.3	1.3	32.8	37.4	18.4	7.2	2.1	average
4	41.7	261.5	14.6	260.8	29.8	1.5	44.9	40.3	21.2	10.5	3.6	average
3	45.6	304.2	35.5	296.3	18.2	1.2	48.4	62.9	22.6	16.8	5.7	average
2	40.6	194.1	10.7	176.4	27	1.4	20.2	28.6	14.1	10.2	1.3	low
1	40.6	216.7	13.3	199.5	25.6	3.8	32.8	37	11.6	10.2	2	average

Table A5-5: Classification results with 10% increased mean interval: low-risk plans and high-risk plans, 1995-98

High risk plans

data 95-98	age	tco	co2	tac	gp	pub	spec	prv	dent	opt	par	risk type
45	66	367.6	78.4	504.3	19	3.5	74.2	116.1	18.6	9.5	5.4	high
40	44.1	443.9	85.2	388.2	13.9	0.3	60.9	51.1	27.1	11.1	9.2	high
38	42.1	376	25.8	285.6	15.1	0.5	58	44.1	30.2	16.6	6.4	high
39	41.9	347.8	41.8	306.2	16.2	0.6	57.9	52.1	32.9	13	4.3	high
30	59	442.3	43.3	456.3	20.3	1.7	76.6	74.4	29.3	21.3	9.8	high
29	51.3	322.9	48.5	327.1	17.3	6	50.8	70.1	19.6	8.4	11.1	high
15	47.8	331.3	69.3	324.3	18.1	1.6	45.6	64.8	19.2	12.1	7.4	high
13	44.9	414.8	45.1	356.8	20	0.5	65.6	63.6	31.3	14.1	5.5	high
8	46.8	413.8	25.9	360.6	22.9	1	59.6	64.8	28.9	15.6	10	high
6	45.6	336.3	31.4	322.7	25.3	0.9	52.1	66.7	23.1	13.7	5.3	high
Total: 10												

Low risk plans

data 95-98	age	tco	co2	tac	gp	pub	spec	prv	dent	opt	par	risk type
43	35.7	114.9	34.4	136	17.9	0.8	14.8	15.7	8.6	3	0.3	low
41	42.9	135.4	18	107.3	18.3	5.7	13.5	14.8	5	4	0.9	low
36	43.9	181.4	14.2	171.1	31.2	1.6	19.8	23.1	10.9	6.5	0.8	low
36	39.7	127.8	12	144.7	26.6	1	14.4	19.7	10.9	8.6	0.7	low
31	39.6	235.5	36.6	215.7	17.4	0.7	29.6	39.6	14.4	8.6	2.2	low
29	37.5	79.2	16.2	103.1	18.9	1.2	9.6	12.2	6.8	4.1	0.3	low
28	41.1	96.4	28.2	134.1	16.3	1.7	13.7	24.9	7.9	7.1	1.1	low
11	37	104.9	20.8	114.2	20.4	1.8	9.1	11.9	7.8	4.4	0.5	low
10	38.4	147.1	11.8	133.9	15.9	1.1	20.3	28.8	6	5.6	0.9	low
Total 10												

Note: Among the high-risk plans for three plans three mean classifications changed: for plan 38 public hospital benefits rose to average, for plan 13 private hospital benefits fell to average, and for plan 6 paramedical services fell to average. None of these mean classifications changed their overall plan classification as high-risk plans. Among the low-risk plans for three plans three mean classifications changed: for plan 43 gp benefits rose to average, for plan 41 contribution rose to average, and for plan 36 claims rose to average. Again, none of these mean classifications changed their overall plan classification as low-risk plans.

Table A5-6: Data plan classification into low- and high-risk plans with 10% decreased mean interval, 1995-98

data 95-98	age	tco	co2	tac	gp	pub	spec	prv	dent	opt	par	risk type
49	41.6	237.3	44.8	251.4	17.7	5	31.5	31.3	19.6	10.6	3.7	average
48	68	367.6	78.4	504.3	19	3.5	74.2	116.1	18.6	9.5	34	high
47	39	256.3	41.8	245.2	17.7	1.9	33.4	50.4	15.4	13.1	2.4	average
46	44.1	443.9	85.2	388.2	13.9	0.3	60.9	51.1	27.1	11.1	9.2	high
45	53.4	297.3	72.8	305.4	5.3	0.3	49	91.1	22.8	5.4	3.3	average
44	42	255.3	45.3	296	17.8	0.3	49.9	77.7	19.9	8.3	4.7	average
43	35.7	114.9	34.4	136	17.9	0.8	14.8	15.7	8.6	3	0.3	low
42	47.4	234.3	19.5	233.4	27.1	2.1	40.1	31.2	17.3	8.6	2.3	average
41	42.9	135.4	18	107.3	18.3	5.7	13.5	14.8	6	4	0.9	low
40	45.9	233.9	15.6	261.9	27.8	1.4	41.2	59.4	16.7	11	3.4	average
39	43.9	181.4	14.2	171.1	31.2	1.6	19.8	23.1	10.9	6.5	0.8	low
38	42.1	376	25.8	285.6	15.1	0.5	55	44.1	30.2	16.6	6.4	high
37	42.8	309.5	60	285.5	14.3	0.7	44.2	56.1	20	7.1	3.8	average
36	39.7	127.8	12	144.7	26.6	1	14.4	19.7	10.9	8.6	0.7	low
35	41.9	347.5	41.5	306.2	16.2	0.6	57.9	52.1	32.9	13	4.3	high
34	46.1	293.6	42.5	270.5	17.3	0.6	39.3	52.8	17.2	8.2	4.9	average
33	37.5	319.3	37.1	269.1	18.4	0.6	57.5	59	18.1	7.3	3	average
32	41.8	269.4	59.8	281.2	14.6	0.8	49.5	68.4	20.3	5.3	1.5	average
31	39.6	235.5	36.9	215.7	17.4	0.7	29.6	39.6	14.4	8.6	2.2	low
30	59	442.3	43.3	456.3	20.3	1.7	76.6	74.4	29.3	21.3	9.8	high
29	37.5	79.2	16.2	103.1	18.9	1.2	9.6	12.2	8.8	4.1	0.3	low
28	41.1	96.4	28.2	134.1	16.3	1.7	13.7	24.9	7.9	7.1	1.1	low
27	34.9	283.9	21.8	221.2	22.3	0.4	36	50.6	18	12.7	2.7	average
26	39.4	302.9	25.9	252.9	18.7	1.3	39.3	48.6	24	12.8	4.6	average
25	51.3	322.9	48.5	327.1	17.3	6	50.8	70.1	19.6	8.4	11.1	high
24	47.2	293.3	45.7	306.6	15.4	0.9	49	71	17	10.4	4.8	average
23	38.3	206.5	40.4	250.1	16.3	0.5	41.2	45.4	17.1	13.1	3.9	average
22	41.5	283.5	39.6	300.5	19.6	2	47.1	65.1	19	8.6	4.1	average
21	39.1	146.9	14.5	180.1	29.6	3.3	21.9	32.3	13.1	8.4	2	average
20	48.1	216.1	52.7	250.6	13.8	2.9	36.8	39.2	18.6	10.6	5.2	average
19	39.4	286.8	17.3	227.5	23	0.5	35	38.6	20.7	11.9	3.2	average
18	47.2	269	50.6	288.3	14.4	1.4	48.1	54.9	17.9	10.9	6.9	average
17	37.4	350.6	28.5	306.9	21.5	0.6	55.7	54.8	29.4	24	5.9	average
16	44.3	331.9	51.7	315.7	16.5	0.9	49.4	61.5	22.3	9.6	8.7	average
15	47.8	331.3	69.3	324.3	18.1	1.8	45.6	64.8	19.2	12.1	7.4	high
14	36.8	245	23.2	207.7	21.8	1.3	33.5	40.3	19.5	9.7	2.2	average
13	44.9	414.8	45.1	356.5	20	0.5	65.6	63.6	31.3	14.1	5.5	high
12	40	241.7	16.6	220.1	24.5	1.4	30.9	38.8	20.9	10.9	3	average
11	37	104.9	20.8	114.2	20.4	1.8	9.1	11.9	7.8	4.4	0.5	low
10	38.4	147.1	11.8	133.9	15.9	1.1	20.3	28.8	6	5.8	0.9	low
9	43.9	271.7	56.3	264.4	15	0.9	40	56.3	18.3	12.6	4.7	average
8	46.8	413.8	25.5	360.8	22.9	1	59.6	64.8	26.9	15.6	10	high
7	40.9	204.7	21	194.7	20.2	1.1	33.8	40	14	7.3	3.3	average
6	45.6	336.3	31.4	322.7	25.3	0.9	52.1	66.7	23.1	13.7	5.3	high
5	36.9	201.7	18.8	188.5	19.3	1.3	32.8	37.4	18.4	7.2	2.1	average
4	41.7	261.5	14.6	260.8	29.8	1.5	44.9	40.3	21.2	10.5	3.6	average
3	45.6	304.2	35.5	296.3	18.2	1.2	48.4	62.9	22.6	16.8	5.7	average
2	40.6	194.1	10.7	176.4	27	1.4	20.2	28.6	14.1	10.2	1.3	low
1	40.6	216.7	13.3	199.5	25.6	3.8	32.8	37	11.6	10.2	2	average

Table A5-7: Classification results with 10% decreased mean interval: low-risk plans and high-risk plans, 1995-98

High risk plans

data 95-98	age	tco	co2	tac	gp	pub	spec	prv	dent	opt	par	risk type
48	68	367.6	78.4	504.1	19	3.6	74.2	116.1	18.6	9.5	34	high
46	44.1	443.9	85.2	388.2	13.9	0.3	60.9	51.1	27.1	11.1	9.2	high
38	42.1	376	25.8	285.6	15.1	0.5	55	44.1	30.2	16.6	6.4	high
38	41.9	347.5	41.5	306.2	16.2	0.6	57.9	52.1	32.9	13	4.3	high
30	59	442.3	43.3	456.3	20.3	1.7	76.6	74.4	29.3	21.3	9.8	high
25	51.3	322.9	48.5	327.1	17.3	6	50.8	70.1	19.6	8.4	11.1	high
18	47.5	331.3	69.3	324.3	18.1	1.6	45.6	64.8	19.2	12.1	7.4	high
13	44.9	414.8	45.1	356.5	20	0.5	65.6	63.6	31.3	14.1	5.5	high
8	46.8	413.8	25.5	360.6	22.9	1	59.6	64.8	28.9	15.8	10	high
6	45.6	336.3	31.4	322.7	25.3	0.8	52.1	66.7	23.1	13.7	5.3	high
Total: 10												

Low risk plans

data 95-98	age	tco	co2	tac	gp	pub	spec	prv	dent	opt	par	risk type
43	35.7	114.9	34.4	136	17.9	0.8	14.6	15.7	8.6	3	0.3	low
41	42.9	135.4	16	107.3	18.3	5.7	13.5	14.8	5	4	0.9	low
39	43.9	181.4	14.2	171.1	31.2	1.6	19.8	23.1	10.9	6.5	0.8	low
36	39.7	127.8	12	144.7	26.6	1	14.4	19.7	10.9	8.6	0.7	low
31	39.6	235.5	36.9	215.7	17.4	0.7	29.6	39.6	14.4	8.6	2.2	low
29	37.5	79.2	16.2	103.1	18.9	1.2	9.6	12.2	8.6	4.1	0.3	low
28	41.1	96.4	28.2	134.1	16.3	1.7	13.7	24.9	7.9	7.1	1.1	low
11	37	104.9	20.8	114.2	20.4	1.8	9.1	11.9	7.6	4.4	0.5	low
10	38.4	147.1	11.8	133.9	15.9	1.1	20.3	28.8	6	5.8	0.9	low
Total 10												

Note: Among the high-risk plans for four plans five mean classifications changed: for plan 35 claims rose to above average and public hospital benefits fell to below average, for plan 25 specialist benefits rose to above average, for plan 13 public hospital benefits fell below average, and for plan 6 contributions rose to above average. None of these mean classification changes altered their overall plan classification, on the contrary, some supported their classification as high-risk plans. Among the low-risk plans for three plans five mean classifications changed: for plan 39 private hospital benefits fell to below average, for plan 28 private hospital and optical benefits fell to below average, and for plan 10 age and contribution feel to below average. All of these mean classification changes rather supported their overall plan classification as low-risk plans.

Appendix 6

Model building process

(1) Variable inclusion verification with univariate model assessment

The results of fitting the univariate logistic regression models to the data are given in table A6-1 on the next page. For each listed variable the table displays the following information: (1) the estimated slope coefficient for the univariable regression model containing only that single variable, (2) the (semi-robust) estimated standard error of the estimated slope coefficient, (3 & 4) z-value and the Wald statistic probability, (5) the 95% confidence interval for the coefficient (6) the Wald test statistic for the hypothesis that the slope coefficient is zero (under the null hypothesis it follows the chi-square distribution with one degree of freedom) and (7) the significance level for the Wald test.

With the exception of the co-payment variable (co2) in Model A, the variables for optical benefits (opt) and co-payment in Model B and some utilisation variables in Model C, where the univariate model failed to converge, the evidence in the table shows some association with the outcome for each variable. Thus upon completion of the univariate analysis, the initial selection of variables for the multivariate analysis could be justified - all univariate models that converged showed a significant Wald statistic.

A familiar problem of this type of univariate analysis is, however, the lack of support for the inclusion of variables that display a weak association with the outcome (see Hosmer and Lemeshow 2000) or, in this case, if univariate models that contain the variable do not converge. In these cases, where numerical problems occurred, the variables were still included because they are scientifically relevant and the data is adequate to support such an analysis. Therefore, on the basis of subject matter considerations, the co-payment variable in Model A, the optical benefit and co-payment variables in Model B and the non-converging variables of Model C were kept.

The initial selection of variables for the multivariate analysis was slightly changed by the transformation of two variables, dependant number and income, into categorical dummies that will assure better interpretations of the results. For both variables the mean category (income=6 and dependants=2) was chosen as the reference category. Variable values below that mean category received a value 1 while those above received a value 2. Next, the model variables are checked for linearity in the logit - one key assumption for logistic regression analysis.

Table A6-1: Univariate logisitc regression for variables of Model A, B and C

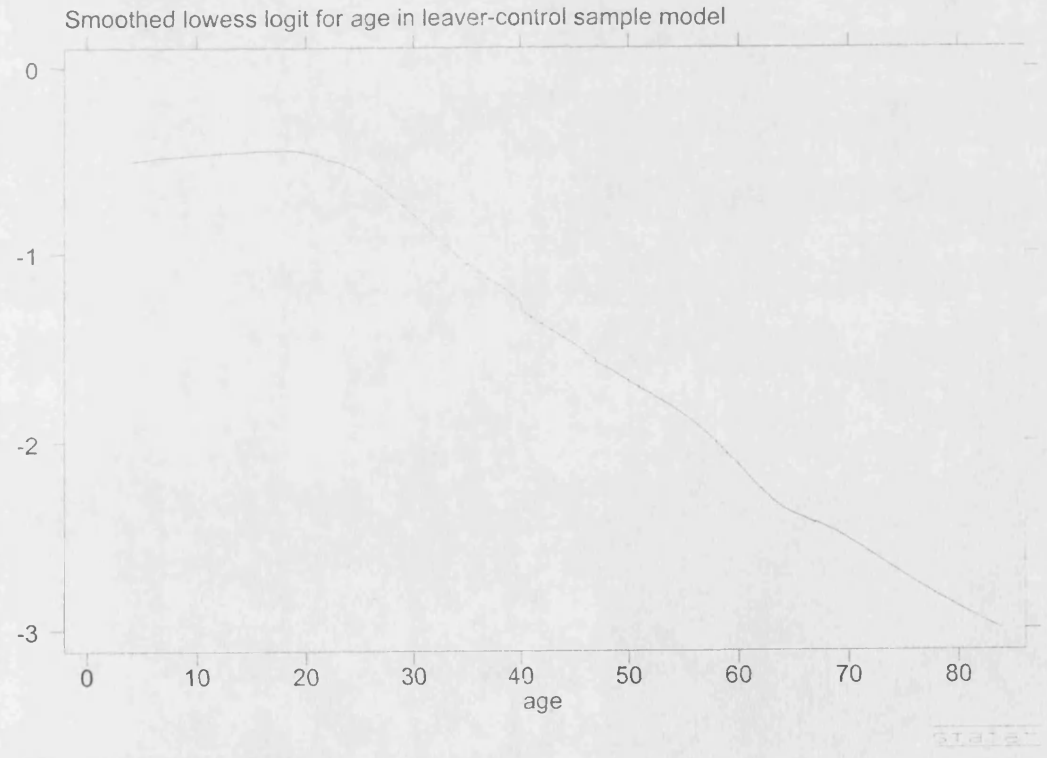
Model A	Coeff.	Std. Err.	z	P>|z|	[95% CI]		Wald chi2(1)	Prob > chi2
variables								
jc								
age	-0.02352**	0.00011	-221.8	0.000	-0.02372	-0.02331	49196.36	0.000
gen	-0.37384**	0.00681	-54.93	0.000	-0.38718	-0.36050	3017.22	0.000
dep	-0.21781**	0.00134	-162.39	0.000	-0.22044	-0.21518	26371.81	0.000
inc	-0.45463**	0.00242	-187.51	0.000	-0.45939	-0.44989	35159.71	0.000
race	-0.00025**	1.63E-06	-151.3	0.000	-0.00025	-0.00024	22892.84	0.000
urb	-0.79237**	0.00448	-176.79	0.000	-0.80116	-0.78359	31255.17	0.000
tac	-0.00303**	3.08E-05	-98.23	0.000	-0.00309	-0.00297	9649.92	0.000
tco	-0.00359**	2.07E-05	-173.84	0.000	-0.00363	-0.00355	30220.9	0.000
co2			estimates diverging					
gp	-0.03431**	0.00024	-142.73	0.000	-0.03478	-0.03384	20372.21	0.000
pub	-0.01824**	0.00187	-9.76	0.000	-0.02190	-0.01458	95.24	0.000
spe	-0.00916**	0.00014	-64.56	0.000	-0.00944	-0.00888	4167.83	0.000
prv	-0.00302**	8.02E-05	-37.65	0.000	-0.00318	-0.00286	1417.89	0.000
den	-0.01454**	0.00021	-68.7	0.000	-0.01495	-0.01412	4719.42	0.000
opt	-0.03040**	0.00036	-85.31	0.000	-0.03110	-0.02970	7278.26	0.000
par	-0.01760**	0.00092	-19.23	0.000	-0.01940	-0.01581	369.64	0.000
risk	-0.37659**	0.00672	-56.08	0.000	-0.38975	-0.36343	3145.4	0.000
N = 663,644								
n = 238,561								
Model B	**Coeff.**	**Std. Err.**	**z**	**P>|z|**	**[95% CI]**		**Wald chi2(1)**	**Prob > chi2**
variables								
lc								
age	-0.01509**	9.38E-05	-160.89	0.000	-0.01527	-0.01491	25886.6	0.000
gen	-0.30279**	0.00646	-46.85	0.000	-0.31546	-0.29013	2195.07	0.000
dep	-0.13348**	0.00114	-116.65	0.000	-0.13572	-0.13123	13607.59	0.000
inc	-0.24723**	0.00203	-122.96	0.000	-0.25370	-0.24574	15117.95	0.000
race	-0.00022**	1.88E-06	-116.11	0.000	-0.00022	-0.00021	13481.97	0.000
urb	-0.53116**	0.00410	-129.6	0.000	-0.53920	-0.52313	16795.25	0.000
tac	-0.00196**	2.72E-05	-71.93	0.000	-0.00201	-0.00190	5173.75	0.000
tco	-0.00335**	1.89E-05	-177.26	0.000	-0.00338	-0.00331	31419.51	0.000
co2			estimates diverging					
gp	-0.02582**	0.00020	-129.17	0.000	-0.02621	-0.02543	16684.65	0.000
pub	-0.00126**	0.00044	-2.87	0.004	-0.00211	-0.00040	8.23	0.000
spe	-0.00626**	0.00013	-48.69	0.000	-0.00651	-0.00601	2370.87	0.000
prv	-0.00152**	0.00008	-18.57	0.000	-0.00168	-0.00136	344.77	0.000
den	-0.00929**	0.00018	-51.48	0.000	-0.00964	-0.00894	2650.69	0.000
opt			estimates diverging					
par	-0.00612**	0.00063	-9.7	0.000	-0.00735	-0.00488	94.07	0.000
risk	-0.27127**	0.00525	-51.68	0.000	-0.28155	-0.26098	2671.28	0.000
N = 684,939								
n = 262,349								
Model C	**Coeff.**	**Std. Err.**	**z**	**P>|z|**	**[95% CI]**		**Wald chi2(1)**	**Prob > chi2**
variables								
sc								
age	-0.05904**	0.00025	-235.3	0.000	-0.05953	-0.05855	55364.33	0.000
gen	-1.99140**	0.02278	-87.430	0.000	-2.03604	-1.94676	7644.620	0.000
dep	-0.86936**	0.00463	-187.9	0.000	-0.87843	-0.86029	177981	0.000
inc	-1.30313**	0.00650	-200.37	0.000	-1.31588	-1.29039	40147.53	0.000
race	-0.00392**	0.00002	-176.5	0.000	-0.00396	-0.00387	31151.58	0.000
urb	-2.47540**	0.00890	-276.86	0.000	-2.49294	-2.45790	76653.400	0.000
tac			estimates diverging					
tco	-0.01084**	0.000068	-160.07	0.000	-0.01097	-0.01071	25621.33	0.000
co2			estimates diverging					
gp	-0.12365**	0.00090	-137.87	0.000	-0.12540	-0.12189	19009.34	0.000
pub			estimates diverging					
spe			estimates diverging					
prv			estimates diverging					
den	-0.08055**	0.00121	-66.6	0.000	-0.08292	-0.07818	4434.95	0.000
opt	-0.13188**	0.00121	-109.14	0.000	-0.13425	-0.12951	11911.25	0.000
par			estimates diverging					
risk	-5.2E-07**	1.92E-08	-27.18	0.000	-5.59E-07	-4.83E-07	738.83	0.000
N = 532,010								
n = 177,981								

*Note: ** significance level of 1%*

(2) Model variable check for linearity in the logit

This modelling process step checks the scale of the continuous model variables or the assumption of linearity in the logit. There are three possible methods for exploring the scale of a continuous variable: (1) a univariate smoothed scatterplot on the logit scale, (2) plots using design variables based on the quartile of distribution, and (3) the method of fractional polynomials. The most accessible standard method is the smoothed scatterplot. The method using the design variable approach is possible, but often leads to inconclusive results because it relies on a limited number of (quartiles) points for the plot (Hosmer and Lemeshow 2000). The fractional polynomials method is very computer intensive and has not yet been developed for models with correlated data. Thus the univariate smoothed scatterplot of the logit against the variables was used to test all variables that entered the models for linearity in the logit.[189] As an example the univariate (lowess) smoothed logit of Model B versus age is displayed in figure A6-1 below, It shows that the relationship is linear in the logit.

Figure A6-1: Univariate (lowess) smoothed logit versus age for Model B



[189] If the results in the scatterplot indicate that the logit is not linear in the covariate then grouping and the use of dummy variables is advised (Hosmer and Lemeshow 2000). Alternatively, higher order terms or other nonlinear scaling in the covariate can be used. However, even in cases where non-linearity in the logit is concluded and variables are transformed into a higher order or used in a nonlinear scaling (ln[x]) they often do not produce better models. In such a case, the variable should be treated as a continuous and linear variable (Hosmer and Lemeshow 2000).

Upon completion of the check of the linearity in the logit the initial selection of variables for the multivariate analysis remains the same. The main effects models are thus specified with the variables shown in table A6-2 on the next pages. Table A6-2 displays the following information for each main effect (listed in the first column): (1) the estimated slope coefficient for the multivariate regression model, (2) the (semi-robust) estimated standard error of the estimated slope coefficient, (3 & 4) z-value and the Wald statistic probability, and (5) the 95% confidence interval for the coefficient. The Wald test statistic and Wald test significance level for the models are shown below each table (under the null hypothesis it follows the chi-square distribution with 20 degrees of freedom).

The results of fitting the multivariate main effects models indicate weaker associations for some covariates when controlling for other variables. In Model A all variables have an association with the output except for the public hospital benefits (pub). For Model B several variables are not significant: above average dependants number (depm_2), above average income (icgm_2), contribution (tco), co-payment (co2), specialist benefits (spe), and private hospital benefits (prv). Model C has five insignificant associations - the covariates dependants number (depm), Asian race (race_1), GP benefits, specialist benefits, and paramedical benefits (par). Strict compliance with the standard levels of statistical significance here would require consideration of smaller models for B and C and deleting the insignificant covariates.

However, all variables in the models have important functions. For instance race and income are essential control variables. The contribution, co-payment, and benefit variables are of key conceptual value and were carefully selected based on theoretical assumptions. They could not be easily dropped or exchanged for different variables without changing the whole nature of the investigation. Further, it would be more sensible, in light of the later result interpretation, to have the same covariates in each model. Thus on conceptual grounds all variables should be kept. Also, at this point the main effects models are not yet controlled for any interactions between the variables, which could prove several of the insignificant associations in fact significant. This will be done next and concerns the assumption of additivity in logistic regression analysis.

Table A6-2: Multivariate main effects models, Models A, B and C

Model A variables lc	Coef.	Std. Err.	z	P>|z|	[95% C I]	
age	-0.05905**	0.0004	-139.15	0.000	-0.05989	-0.05822
gen	0.58935**	0.0101	58.35	0.000	0.56956	0.60915
_idepm_1	-0.01685**	0.0048	-3.53	0.000	-0.02621	-0.00750
_idepm_2	-0.02710**	0.0047	-5.78	0.000	-0.03629	-0.01791
_icgm_1	0.03018**	0.0040	7.53	0.000	0.02232	0.03803
_icgm_2	-0.01802**	0.0039	-4.58	0.000	-0.02574	-0.01030
_irace_1	-0.82800**	0.0185	-44.65	0.000	-0.86435	-0.79165
_irace_2	-0.59055**	0.0152	-38.87	0.000	-0.62033	-0.56078
_irace_3	-1.41969**	0.0135	-105	0.000	-1.44619	-1.39319
urb	0.15489**	0.0148	10.47	0.000	0.12591	0.18388
tac	-0.00013**	0.0000	-11.52	0.000	-0.00015	-0.00011
tco	0.00058**	0.0000	39.02	0.000	0.00055	0.00060
co2	0.00019**	0.0000	12.83	0.000	0.00016	0.00022
gp	0.00024**	0.0000	6.27	0.000	0.00017	0.00032
pub	0.00004	0.0000	1.29	0.197	-0.00002	0.00009
spe	0.00018**	0.0000	11.97	0.000	0.00015	0.00021
prv	0.00014**	0.0000	11.45	0.000	0.00012	0.00016
den	0.00027**	0.0000	14.87	0.000	0.00023	0.00030
opt	0.00038**	0.0000	14.97	0.000	0.00033	0.00043
par	0.00015**	0.0000	6.03	0.000	0.00010	0.00021
_irisk_1	0.45164**	0.0151	29.91	0.000	0.42205	0.48123
_irisk_2	-0.13188**	0.0172	-7.65	0.000	-0.16568	-0.09808
_cons	1.52849**	0.0239	63.95	0.000	1.48164	1.57534
Wald chi2(20) = 34125.25						N = 663,644
Prob > chi2 = 0.000						n = 238,561
Model B variables lc	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**[95% C I]**	
age	-0.03321**	0.00034	-98.58	0.000	-0.03387	-0.03255
gen	0.20322**	0.00892	22.78	0.000	0.18573	0.22071
_idepm_1	-0.04768**	0.00482	-9.89	0.000	-0.05713	-0.03823
_idepm_2	0.00665	0.00464	1.43	0.153	-0.00246	0.01575
_icgm_1	0.02185**	0.00414	5.27	0.000	0.01373	0.02997
_icgm_2	-0.00726	0.00412	-1.76	0.078	-0.01533	0.00081
_irace_1	-0.69824**	0.01725	-40.49	0.000	-0.73205	-0.66444
_irace_2	-0.51034**	0.01408	-36.25	0.000	-0.53793	-0.48275
_irace_3	-0.38247**	0.01078	-35.47	0.000	-0.40360	-0.36133
urb	0.14571**	0.01387	10.5	0.000	0.11852	0.17290
tac	0.00004*	0.00002	2.18	0.029	0.00000	0.00007
tco	0.00002	0.00001	1.55	0.122	-0.00001	0.00005
co2	-0.00001	0.00002	-0.37	0.712	-0.00004	0.00003
gp	0.00097**	0.00005	20.65	0.000	0.00088	0.00107
pub	0.00010**	0.00003	3.12	0.002	0.00004	0.00016
spe	-0.00002	0.00002	-0.88	0.377	-0.00006	0.00002
prv	-0.00003	0.00002	-1.56	0.119	-0.00006	0.00001
den	0.00037**	0.00002	15.43	0.000	0.00033	0.00042
opt	0.00111**	0.00003	35.23	0.000	0.00105	0.00117
par	0.00008*	0.00003	2.52	0.012	0.00002	0.00015
_irisk_1	-0.03105*	0.01427	-2.18	0.030	-0.05901	-0.00308
_irisk_2	0.29832**	0.01406	21.21	0.000	0.27076	0.32589
_cons	0.84594**	0.02137	39.58	0.000	0.80404	0.88783
Wald chi2(20) = 17332.18						N = 684,939
Prob > chi2 = 0.000						n = 262,349

Cont.

Cont. Model C variables sc	Coef.	Std. Err.	z	P>|z|	[95% C I]
age	-0.00801**	0.00066	-12.16	0.000	-0.00930 -0.00672
gen	0.01086	0.02053	0.53	0.597	-0.02939 0.05110
_idepm_1	-0.00406	0.01184	-0.34	0.732	-0.02727 0.01915
_idepm_2	-0.04411**	0.01164	-3.79	0.000	-0.06691 -0.02130
_icgm_1	0.05577**	0.01093	5.1	0.000	0.03436 0.07719
_icgm_2	-0.09842**	0.01081	-9.11	0.000	-0.11960 -0.07724
_irace_1	-0.04536	0.02952	-1.54	0.124	-0.10322 0.01250
_irace_2	-0.92967**	0.03812	-24.39	0.000	-1.00438 -0.85496
_irace_3	-1.67975**	0.02994	-56.11	0.000	-1.73843 -1.62107
urb	0.28379**	0.03454	8.22	0.000	0.21610 0.35148
tac	-0.00005*	0.00002	-2.15	0.032	-0.00009 0.00000
tco	0.00021**	0.00004	5.79	0.000	0.00014 0.00028
co2	0.00030**	0.00003	8.69	0.000	0.00023 0.00036
gp	0.00002	0.00012	0.13	0.894	-0.00021 0.00024
pub	-0.00017**	0.00006	-3.01	0.003	-0.00029 -0.00006
spe	0.00005	0.00003	1.48	0.138	-0.00001 0.00011
prv	0.00006*	0.00002	2.48	0.013	0.00001 0.00011
den	0.00021**	0.00005	4.2	0.000	0.00011 0.00030
opt	-0.00034**	0.00009	-3.87	0.000	-0.00051 -0.00017
par	-0.00012	0.00006	-1.87	0.061	-0.00024 0.00001
_irisk_1	0.64090**	0.03880	16.52	0.000	0.56486 0.71694
_irisk_2	-0.06705**	0.02662	-2.52	0.012	-0.11923 -0.01487
_cons	-2.00664**	0.04921	-40.77	0.000	-2.10310 -1.91019
Wald chi2(20) = 4729.87					N = 532,010
Prob > chi2 = 0.000					n = 177,981

*Note: ** significance level of 1% and * significance level of 5%*

(3) Model check for interactions

This model building step assesses the need to include interactions. Whether or not to include an interaction term in the model is mostly dependent on practical considerations, and not only on statistical concerns. This is particularly true if interactions can lead to practical, relevant interpretations. First, a list that covers all possible interaction terms for a model should be prepared. Of this list of possible interactions, scientifically plausible interactions should be selected and then each interaction is added one at a time to the model to be tested for significance. It is important to note that not all possible interactions have to be included in this step but only scientifically plausible interactions (Hosmer and Lemeshow 2000).

The list of scientifically plausible interaction terms for the three Models A, B, and C is displayed in the first columns of tables A6-4 to A6-6 at the end of this section. The main effects models contain 17 variables and there are 136 possible pairwise interactions. All interaction terms for Models A, B, and C thought to offer a scientifically plausible modification of the covariate effects. They were added one at a time to the main effects

models and tested with the Wald test (presented in tables A6-4 to A6-6).[190] The results show that several interactions are significant at the 1% level, 54 for Model A, 60 for Model B, and 44 for Model C.

Finally, models with all significantly tested interaction terms are estimated and the final model is selected according to the improvement in fit and results interpretation. If an interaction term is not significant or does not increase the model's significance it should not be included (Hosmer and Lemeshow 2000).[191] Interaction terms that were significant for Model A, B, and C were added to the main effects models. However, these models did not converge. Next, step-by-step significant interaction terms were excluded depending on whether they either did not improve the fit of the model (i.e. the Wald statistic did not increase) or prevented the model from converging to a solution.[192] This complicated process yielded the preliminary final models.

Table A6-3 on the next page represents the fitted models containing the main effects and the significant interaction terms. The table displays the following information for each main effect and interaction listed: (1) the estimated slope coefficient for the multivariate regression model, (2) the (semi-robust) estimated standard error of the estimated slope coefficient, (3 & 4) z-value and the Wald statistic probability, and (5) the 95% confidence interval for the coefficient. The Wald test statistic and Wald test significance level for the models are shown below each table (under the null hypothesis it follows the chi-square distribution with 37 degrees of freedom). For all three models the Wald statistic increased with the inclusion of the interaction terms. All three preliminary final models A, B, and C have 17 main variables and 15 interactions.

In preliminary final Model A and B all variables and interactions are significant. However, Model C still contains three insignificant variables: above average dependant number, above average income and paramedical benefits and one insignificant interaction race*special benefits. These variables were retained because they function as important control variables (and the below average dependant and income variables are significant).

[190] Usually not all possible interaction terms are scientifically plausible, however, in this case all 136 were practically likely and thus had to be tested for significance.

[191] An interaction term can only alter point and interval estimates if its estimated coefficient is statistically significant.

[192] The results for each step of this stepwise approach are not presented here, however, the results are available from the author upon request. Best fitted converging models (also not presented here) that included the maximum of the significant interaction terms was the result of this stepwise elimination approach. However, these models revealed Wald statistics for some of the included interaction terms that were not significant and they were dropped subsequently. Also, for each model three interaction terms (age*income, race*income, gender*income, and urban*income) that were first eliminated by the stepwise approach were reinstated in the final model for conceptual reasons, if they yielded significant Wald statistics.

The interaction race*special benefit ensured the significance of the special benefit covariate and was therefore not excluded. Finally, the insignificant coefficient for the paramedical benefits variable has interpretational value as a result.

Table A6-3: Preliminary final models with main effects and interaction, Models A, B and C

Model A variables |c	Coef.	Std. Err.	z	P>|z|	[95% C I]
age	-0.06188**	.0005983	-103.44	0.000	-0.06305 -0.06071
gen	0.13556**	0.03254	4.17	0.000	0.07179 0.19933
_idepm_1	-0.02544**	0.00532	-4.78	0.000	-0.03587 -0.01501
_idepm_2	-0.03039**	0.00621	-4.89	0.000	-0.04256 -0.01822
_icgm_1	0.11218**	0.01209	9.28	0.000	0.08848 0.13589
_icgm_2	0.13465**	0.02300	5.86	0.000	0.08958 0.17972
_irace_1	-0.82924**	0.01863	-44.52	0.000	-0.86575 -0.79273
_irace_2	-0.61434**	0.01670	-36.78	0.000	-0.64707 -0.58160
_irace_3	-1.42872**	.0138872	-102.88	0.000	-1.45594 -1.40150
urb	0.22973**	0.01882	12.21	0.000	0.19284 0.26662
tac	-0.00013**	0.00001	-11.37	0.000	-0.00015 -0.00011
tco	0.00054**	0.00002	31.62	0.000	0.00051 0.00058
co2	0.00019**	0.00001	12.63	0.000	0.00016 0.00021
gp	0.00045**	0.00013	3.48	0.001	0.00020 0.00070
pub	-0.00017*	0.00007	-2.35	0.019	-0.00031 -0.00003
spe	0.00018**	0.00001	11.84	0.000	0.00015 0.00021
prv	0.00013**	0.00001	11.12	0.000	0.00011 0.00016
den	0.00060**	0.00006	9.34	0.000	0.00047 0.00073
opt	0.00086**	0.00007	12.15	0.000	0.00072 0.00100
par	0.00016**	0.00003	5.75	0.000	0.00010 0.00021
_irisk_1	0.31103**	0.02147	14.49	0.000	0.26895 0.35311
_irisk_2	-0.33860**	0.02411	-14.05	0.000	-0.38584 -0.29135
age*gen	0.01026**	0.00082	12.59	0.000	0.00866 0.01186
age*icgm	-0.00079**	0.00021	-3.82	0.000	-0.00120 -0.00039
age*depm	-0.00001**	0.00000	-4.77	0.000	-0.00001 0.00000
age*opt	-0.00001**	0.00000	-7.45	0.000	-0.00002 -0.00001
age*pub	0.00001**	0.00000	3.22	0.001	0.00000 0.00001
depm*gen	0.04221**	0.00510	8.28	0.000	0.03221 0.05221
depm*gp	-0.00037**	0.00006	-6.14	0.000	-0.00048 -0.00025
icgm*race	-0.06063**	0.00869	-6.98	0.000	-0.07766 -0.04360
icgm*den	-0.00015**	0.00002	-6.05	0.000	-0.00020 -0.00010
race*gen	-0.00024**	0.00003	-7.77	0.000	-0.00029 -0.00018
race*den	0.00013**	0.00001	9.52	0.000	0.00010 0.00015
gp*tco	0.00000**	0.00000	6.1	0.000	0.00000 0.00000
risk*gen	0.06285**	0.01559	4.03	0.000	0.03229 0.09341
risk*icgm	0.04410**	0.00352	12.54	0.000	0.03721 0.05100
risk*race	0.04174**	0.00913	4.57	0.000	0.02385 0.05963
_cons	1.60846**	0.03091	52.04	0.000	1.54788 1.66904
Wald chi2(37)	= 35821.34				N = 663,644
Prob > chi2	= 0.000				n = 238,561

Cont.

Cont.

Model B variables Ic	Coef.	Std. Err.	z	P>|z|	[95% C I]
age	-0.03612**	0.00064	-56.08	0.000	-0.03738 -0.03485
gen	0.68371**	0.02963	23.07	0.000	0.62563 0.74179
_idepm_1	-0.06332**	0.00948	-6.68	0.000	-0.08191 -0.04473
_idepm_2	-0.05722**	0.01702	-3.36	0.001	-0.09057 -0.02386
_icgm_1	0.16825**	0.01195	14.07	0.000	0.14482 0.19168
_icgm_2	0.22438**	0.02366	9.48	0.000	0.17800 0.27076
_irace_1	-0.99478**	0.02053	-48.45	0.000	-1.03503 -0.95454
_irace_2	-1.10995**	0.02573	-43.14	0.000	-1.16038 -1.05952
_irace_3	-1.27728**	0.03342	-38.22	0.000	-1.34278 -1.21178
urb	0.21248**	0.01809	11.75	0.000	0.17702 0.24793
tac	0.00005**	0.00001	3.9	0.000	0.00003 0.00008
tco	0.00024**	0.00002	10	0.000	0.00019 0.00029
co2	0.00027**	0.00005	5.59	0.000	0.00018 0.00036
gp	0.00093**	0.00004	22.22	0.000	0.00085 0.00101
pub	-0.00015*	0.00007	-2.01	0.045	-0.00029 0.00000
spe	-0.00006**	0.00002	-3.73	0.000	-0.00010 -0.00003
prv	-0.00008**	0.00001	-5.95	0.000	-0.00010 -0.00005
den	0.00111**	0.00006	19.75	0.000	0.00100 0.00122
opt	0.00198**	0.00008	23.86	0.000	0.00182 0.00214
par	0.00008*	0.00004	2.23	0.026	0.00001 0.00015
_irisk_1	-0.08470**	0.01501	-5.64	0.000	-0.11412 -0.05528
_irisk_2	0.21934**	0.01864	11.77	0.000	0.18281 0.25587
age*gen	-0.01187**	0.00071	-16.83	0.000	-0.01325 -0.01049
age*depm	0.00067**	0.00021	3.21	0.001	0.00026 0.00107
age*icgm	-0.00250**	0.00019	-13.22	0.000	-0.00287 -0.00213
age*race	0.00759**	0.00026	29.63	0.000	0.00709 0.00810
age*den	-0.00002**	0.00000	-14.67	0.000	-0.00002 -0.00002
age*opt	-0.00002**	0.00000	-11.91	0.000	-0.00002 -0.00002
age*co2	-0.00001**	0.00000	-6.19	0.000	-0.00001 0.00000
age*pub	0.00000*	0.00000	2.5	0.012	0.00000 0.00001
age*tac	0.00000**	0.00000	2.88	0.004	0.00000 0.00000
gen*icgm	-0.01372**	0.00471	-2.91	0.004	-0.02294 -0.00449
race*tco	-0.00019**	0.00001	-17.63	0.000	-0.00021 -0.00017
icgm*race	0.02922**	0.00206	14.17	0.000	0.02518 0.03326
icgm*urb	-0.05555**	0.00889	-6.25	0.000	-0.07298 -0.03812
opt*tac	0.00000**	0.00000	-2.85	0.004	0.00000 0.00000
risk*icgm	0.03389**	0.00372	9.11	0.000	0.02659 0.04118
_cons	0.87973**	0.03334	26.38	0.000	0.81438 0.94508
Wald chi2(37)	= 20066.5				N = 684,939
Prob > chi2	= 0.000				n = 262,349

Cont.

Cont.

Model C variables sc	Coef.	Std. Err.	z	P>|z|	[95% C I]
age	-0.00671**	0.00101	-6.66	0.000	-0.00869 -0.00474
gen	0.24360**	0.06394	3.81	0.000	0.11828 0.36891
_idepm_1	0.01540	0.01187	1.3	0.195	-0.00787 0.03866
_idepm_2	-0.05222**	0.01166	-4.48	0.000	-0.07508 -0.02936
_icgm_1	0.08251**	0.01188	6.94	0.000	0.05922 0.10580
_icgm_2	0.00203	0.01718	0.12	0.906	-0.03165 0.03570
_irace_1	0.13200**	0.03299	4	0.000	0.06735 0.19665
_irace_2	-0.65239**	0.04349	-15	0.000	-0.73763 -0.56715
_irace_3	-1.27134**	0.04804	-26.46	0.000	-1.36550 -1.17719
urb	0.11245*	0.04903	2.29	0.022	0.01636 0.20854
tac	-0.00007**	0.00002	-2.93	0.003	-0.00012 -0.00002
tco	-0.00063**	0.00013	-4.96	0.000	-0.00088 -0.00038
co2	0.00031**	0.00003	8.83	0.000	0.00024 0.00037
gp	0.00103**	0.00027	3.8	0.000	0.00050 0.00157
pub	-0.00085**	0.00017	-4.85	0.000	-0.00119 -0.00050
spe	0.00008*	0.00003	2.28	0.022	0.00001 0.00014
prv	0.00008*	0.00003	3.21	0.001	0.00003 0.00013
den	0.00022**	0.00005	4.36	0.000	0.00012 0.00032
opt	-0.00077**	0.00011	-6.98	0.000	-0.00098 -0.00055
par	-0.00008	0.00006	-1.28	0.200	-0.00021 0.00004
_irisk_1	0.42828**	0.08650	4.95	0.000	0.25875 0.59782
_irisk_2	-0.56429**	0.17065	-3.31	0.001	-0.89875 -0.22983
age*gen	-0.00987**	0.00139	-7.12	0.000	-0.01258 -0.00715
age*tco	0.00002**	0.00000	6.14	0.000	0.00001 0.00002
gen*icgm	0.04919**	0.01299	3.79	0.000	0.02372 0.07466
gen*gp	0.00093**	0.00022	4.22	0.000	0.00050 0.00137
icgm*race	-0.08189**	0.00586	-13.98	0.000	-0.09337 -0.07041
icgm*gp	-0.00085**	0.00017	-5.05	0.000	-0.00118 -0.00052
race*spe	-0.00004	0.00002	-1.81	0.071	-0.00009 0.00000
race*tco	-0.00015**	0.00003	-4.38	0.000	-0.00022 -0.00008
urb*pub	0.00069**	0.00018	3.78	0.000	0.00033 0.00104
risk*age	-0.00601**	0.00086	-6.97	0.000	-0.00770 -0.00432
risk*gen	0.25842**	0.02626	9.84	0.000	0.20697 0.30988
risk*icgm	0.02203*	0.00894	2.46	0.014	0.00451 0.03955
risk*urb	0.31010**	0.07107	4.36	0.000	0.17080 0.44940
risk*opt	0.00079**	0.00010	7.73	0.000	0.00059 0.00099
risk*tco	0.00031**	0.00005	6.81	0.000	0.00022 0.00041
_cons	-1.88824**	0.07114	-26.54	0.000	-2.02767 -1.74882
Wald chi2(37)	= 5431.19				N = 532,010
Prob > chi2	= 0.000				n = 177,981

*Note: ** significance level of 1% and * significance level of 5%*

The models in table A6-3 are referred to as preliminary final models because they are not yet checked for model adequacy. The next and final step in the model building process will assess the model fit. This step is presented after the result interpretation in chapter five.

Table A6-4: List of plausible interaction terms for Model A with p-value and Wald test statistic when added to the main effects model

age	age															
	[redacted]															
depm	0.000	depm														
	32914.2	[redacted]														
gen	0.000	0.000	gen													
	33190.5	33023.6	[redacted]													
icgm	0.000	0.640	0.000	icgm												
	32915.0	32915.65	32935.7	[redacted]												
race	0.003	0.002	0.000	0.087	race											
	33189.0	32944.4	32892.8	32990.9	[redacted]											
urb	0.444	0.067	0.177	0.000	0.000	urb										
	32917.92	32916.25	32918.0	32956.1	32929.8	[redacted]										
gp	0.461	0.000	0.026	0.000	0.000	0.003	gp									
	32952.85	33074.81	32995.92	32928.71	33093.35	32933.7	[redacted]									
spe	0.069	0.751	0.000	0.000	0.845	0.000	0.022	spe								
	32922.34	32915.47	32936.74	32930.21	32914.21	32923.1	32927.92	[redacted]								
prv	0.743	0.220	0.023	0.646	0.063	0.838	0.012	diverging	prv							
	32913.09	32917.09	32910.14	32918.72	32918.88	32914.13	32933.49	diverging	[redacted]							
pub	0.000	0.537	0.872	0.443	0.332	0.464	0.030	0.001	0.124	pub						
	32934	32915.53	32914.65	32915.54	32913.87	32917.36	32916.55	32915.07	32915.3	[redacted]						
den	0.000	0.382	0.000	0.000	0.000	0.095	0.000	0.405	0.095	0.349	den					
	33019.14	32914.59	32930.5	32970.13	33017.77	32916.98	32954.5	32915.6	32923.27	32918.02	[redacted]					
opt	0.000	0.635	0.070	0.000	0.000	0.094	0.259	0.005	0.081	0.044	0.062	opt				
	32992.19	32914.87	32917.52	32924.64	32925.85	32915.56	32921.8	32917.15	32915.81	32915.29	32948.26	[redacted]				
par	0.621	0.077	0.004	0.027	0.122	0.145	0.005	0.317	diverging	0.289	diverging	0.000	par			
	32917.37	32919.44	32922.84	32942.25	32915.94	32916.41	32931.34	32929.45	diverging	32920.87	diverging	32924.02	[redacted]			
tac	0.300	0.054	0.000	0.000	0.982	0.222	0.000	diverging	diverging	0.000	0.000	0.095	diverging	tac		
	32921.06	32917.64	32932.84	32932.19	32915.81	32917.63	32942.41	diverging	diverging	32910.77	32933.19	32903.97	diverging	[redacted]		
tco	0.000	0.000	0.008	0.001	0.000	0.031	0.000	0.732	0.735	0.282	0.000	0.028	0.451	0.229	tco	
	32950.48	32919.87	32927.53	32911.8	32848.88	32921.97	33148.39	32929.14	32918.97	32913.55	32914.36	32947.38	32919.06	32937.63	[redacted]	
co2	0.000	0.066	0.447	0.825	0.000	0.535	0.234	diverging	diverging	0.011	0.000	0.006	0.000	0.986	0.623	co2
	32951.09	32930.8	32924.15	32914.46	32883.01	32918.82	32915.05	diverging	diverging	32913.1	32922.46	32887.57	32945.55	32972.5	32911.21	[redacted]
risk	0.000	0.136	0.000	0.000	0.000	0.000	0.000	0.018	0.265	0.442	0.558	0.000	0.965	0.137	0.171	0.000
	33900.52	34141.47	34171.7	34367.58	34230.72	34146.59	34139.22	34126.78	34127.95	34126.31	34125.66	34139.42	34126.45	34128.22	34124.89	34119.56

Note: Each cell contains two values, first the p-value of the significance test when added alone to the main effects model and below the Wald statistic of the new model as a measure of fit. The Wald statistic for the main effects model was 32914.37.

Table A6-5: List of plausible interaction terms for Model B with p-value and Wald test statistic when added to the main effects model

age	age															
depm	0.000	depm														
	16830.5															
gen	0.000	0.129	gen													
	17073.21	16786.58														
icgm	0.000	0.000	0.000	icgm												
	16956.95	16794.71	16807.06													
race	0.000	0.001	0.727	0.000	race											
	17291.18	16788.11	16784.87	16891.98												
urb	0.004	0.932	0.000	0.000	0.112	urb										
	16799.93	16784.07	16812.73	16841.11	16795.09											
gp	0.014	0.150	0.024	0.015	0.000	0.000	gp									
	16749.33	16795.82	16787.53	16781.05	16795.34	16795.84										
spe	0.000	0.152	0.741	0.000	0.205	0.001	0.979	spe								
	16866.41	16790.14	16783.53	16792.25	16779.76	16795.96	16815.36									
prv	0.095	0.689	0.194	0.342	0.035	0.272	0.535	diverging	prv							
	16777.95	16785.58	16785.03	16785.26	16776.16	16789.12	16798.88	diverging								
pub	0.003	0.447	0.148	0.791	0.045	0.817	0.254	0.183	0.452	pub						
	16769.65	16784.69	16790.27	16794.8	16789.65	16787.29	16791.87	16784.25	16786.08							
den	0.000	0.000	0.000	0.001	0.000	0.166	0.000	0.000	0.003	0.001	den					
	17177.04	16815.79	16791.95	16768.58	16774.53	16784.77	16812.67	16795.25	16786.06	16783.32						
opt	0.000	0.000	0.000	0.306	0.011	0.285	0.000	0.000	0.002	0.246	0.000	opt				
	17058.83	16819.51	16749.71	16793.53	16783.63	16784.36	16845.21	16789.58	16790.43	16786.41	16990.89					
par	0.113	0.118	0.556	0.003	0.061	0.259	0.643	0.116	diverging	0.926	0.146	0.695	par			
	16774.22	16787.19	16788.41	16771.74	16785	16784.44	16825.69	16819.38	diverging	16784.83	16783.88	16784.08				
tac	0.000	0.115	0.209	0.056	0.250	0.012	0.895	0.000	diverging	0.083	0.460	0.002	diverging	tac		
	16935.28	16784.78	16786.05	16784.65	16784.55	16792.98	16818.43	16917.39	diverging	16825.86	16781.8	16981.71	diverging			
tco	0.000	0.134	0.000	0.021	0.000	0.892	0.001	0.445	0.150	0.423	0.000	0.000	0.059	0.013	tco	
	15757.27	16789.81	16902.52	16784.43	17780.38	16784.95	16799.65	16786.69	16785.4	16792.46	16808	16851.65	16775.73	16796.47		
co2	0.000	0.008	0.857	0.135	0.634	0.001	0.189	0.005	0.000	0.000	0.063	0.870	diverging	0.000	0.010	co2
	16919.26	16793.23	16784.06	16809.31	16792.78	16808.88	16817.43	16821.1	16814.51	16825.44	16768.35	16800.9	diverging	16847.75	16780.42	
risk	0.000	0.058	0.786	0.000	0.000	0.000	0.000	0.319	0.854	0.004	0.000	0.000	0.119	0.501	0.001	0.453
	17379.43	17335.43	17335.45	17426.84	17350.05	17407.69	17343.06	17333.06	17334.32	17328.61	17319.11	17341.68	17328.86	17332.18	17335.83	17332.86

Note: Each cell contains two values, first the p-value of the significance test when added alone to the main effects model and below the Wald statistic of the new model as a measure of fit. The Wald statistic for the main effects model was 16784.07.

Table A6-6: List of plausible interaction terms for Model C with p-value and Wald test statistic when added to the main effects model

age	age														
depm	0.754	depm													
	4211.86														
gen	0.000	0.117	gen												
	4263.5	4221.7													
icgm	0.005	0.000	0.000	icgm											
	4240.18	4227.57	4214.92												
race	0.087	0.000	0.138	0.000	race										
	4205.82	4239.19	4218.74	4273.75											
urb	0.130	0.000	0.224	0.416	0.007	urb									
	4217.94	4230.18	4212.79	4215.62	4186.68										
gp	0.043	0.010	0.000	0.000	0.356	0.133	gp								
	4214.19	4216.98	4245.83	4250.56	4214.32	4220.29									
spe	0.645	0.140	0.72	0.885	0.009	0.037	0.481	spe							
	4221.34	4216.84	4213.13	4213.93	4217.79	4216.58	4213.41								
prv	0.385	0.375	0.373	0.645	0.043	0.162	0.908	0.188	prv						
	4217.82	4212.49	4213.68	4214.36	4213.77	4212.07	4212.88	4212.22							
pub	0.147	0.264	0.4	0.306	0.000	0.001	0.060	0.340	0.032	pub					
	4214.71	4212.82	4216.46	4213.67	4238.29	4236.8	4213.52	4279.09	4232.36						
den	0.000	0.380	0.525	0.685	0.023	0.297	0.105	0.468	0.952	0.242	den				
	4217.31	4212.97	4212.82	4220.04	4214.25	4214.12	4210.71	4213.16	4212.35	4228.37					
opt	0.000	0.526	0.207	0.557	0.225	0.023	0.173	0.440	0.986	0.150	0.118	opt			
	4236.9	4211.84	4214.78	4212.61	4216.9	4217.23	4212.4	4212.13	4211.92	4212.52	4214.09				
par	0.890	0.000	0.405	0.425	0.646	0.513	0.739	0.917	0.559	0.004	0.612	0.487	par		
	4212.36	4224.69	4224.58	4211.68	4210.92	4215.13	4220.2	4214.68	4218.08	4213.03	4212.53	4215.37			
tac	0.127	0.032	0.393	0.742	0.051	0.414	0.816	0.155	0.431	0.081	0.010	0.000	0.056	tac	
	4205.73	4213	4216.27	4215.01	4209.49	4214.54	4214.23	4211.48	4201.6	4211.78	4213.71	4208.19	4249.99		
tco	0.000	0.026	0.513	0.000	0.000	0.977	0.310	0.000	0.000	0.139	0.202	0.000	0.007	0.000	tco
	4336.75	4235.9	4212.89	4457.96	4364.08	4212.11	4218.63	4258.21	4223.07	4216.49	4217.75	4261.54	4220.21	4267.09	
co2	0.024	0.005	0.081	0.089	0.001	0.332	0.006	0.241	0.000	0.310	0.211	0.117	0.000	0.133	0.312
	4201.64	4226.31	4217.4	4226.25	4215.95	4219.42	4197.41	4215	3974.85	4214.91	4215.51	4192.12	4287.92	4241.76	4227.97
risk	0.000	0.008	0.000	0.027	diverging	0.000	0.000	0.000	0.643	0.256	0.000	0.000	0.048	0.000	0.000
	4792.64	4753.14	5053.08	4727.55	diverging	4781.71	4782.67	4742.51	4730.08	4736.22	4737.23	4811.36	4739.73	4753.33	4800.79

	0.072 risk
	4743.28

Note: Each cell contains two values, first the p-value of the significance test when added alone to the main effects model and below the Wald statistic of the new model as a measure of fit. The Wald statistic for the main effects model was 4211.93

Appendix 7

Calculated odds ratios for models' interactions

Table A7-1: Estimated odds ratios for gender for Model A, controlling for dependants

gender*	dep1	dep2	dep3	dep4	dep5	dep6	dep7	dep8
OR (gen = 1)	1.10	1.25	1.30	1.36	1.41	1.48	1.54	1.61

*Note: * Reference category for gender is male.*

Table A7-2: Estimated odds ratios for income for Model A, controlling for risk

risk*	income**	Odds ratio
risk = 1 (below)	income = 1 (below)	1.17
	income = 2 (above)	1.20
risk = 2 (above)	income = 1 (below)	1.22
	income = 2 (above)	1.25

*Note: * Reference category for risk (plan) is average-risk plan,*
*** Reference category for income is average income group.*

Table A7-3: Estimated odds ratios for risk for Model A, controlling for income

income*	risk**	Odds ratio
income = 1 (below)	risk = 1 (below)	1.43
	risk = 2 (above)	0.75
income = 2 (above)	risk = 1 (below)	1.49
	risk = 1 (above)	0.78

*Note: * Reference category for income is average income group,*
*** Reference category for risk (plan) is average-risk plan.*

Table A7-4: Estimated odds ratios for gender for Model B, controlling for age

gender*	age20	age30	age40	age50	age60	age70	age80
OR (gen = 1)	1.56	1.38	1.23	1.09	0.97	0.86	0.77

*Note: * Reference category for gender is male.*

Table A7-5: Estimated odds ratios for dependants for Model B, controlling for age

dependants*		age20	age40	age60	age80
OR dependants	(below = 1)	0.95	0.96	0.98	0.99
	(above = 2)	0.96	0.97	0.98	1.00

*Note: * Reference category for dependants is average dependants number.*

Table A7-6: Estimated odds ratios for income for Model B, controlling for age

income*		age20	age40	age60	age80
OR income	(below = 1)	1.13	1.07	1.02	0.97
	(above = 2)	1.19	1.13	1.08	1.03

*Note: * Reference category for income is average income group.*

Table A7-7: Estimated odds ratios for income for Model B, controlling for risk

risk*	income**	Odds ratio
risk = 1 (below)	income = 1 (below)	1.22
	income = 2 (above)	1.30
risk = 2 (above)	income = 1 (below)	1.27
	income = 2 (above)	1.34

*Note: * Reference category for risk (plan) is average-risk plan,*
*** Reference category for income is average income group.*

Table A7-8: Estimated odds ratios for urban for Model B, controlling for income

income*	urban**	Odds ratio
income = 1 (below)	urban = 1	1.17
income = 2 (above)	urban = 1	1.11

*Note: * Reference category for income is average income group,*
*** Reference category for urban is rural residency.*

Table A7-9: Estimated odds ratios for risk for Model B, controlling for income

income*	risk**	Odds ratio
income = 1 (below)	risk = 1 (below)	0.95
	risk = 2 (above)	1.29
income = 2 (above)	risk = 1 (below)	0.98
	risk = 1 (above)	1.33

*Note: * Reference category for income is average income group,*
*** Reference category for risk (plan) is average-risk plan.*

Table A7-10: Estimated odds ratios for gender for Model C, controlling for age

gender*	age20	age30	age40	age50	age60	age70	age80
OR (gen = 1)	1.05	0.95	0.86	0.78	0.71	0.64	0.58

*Note: * Reference category for gender is male.*

Table A7-11: Estimated odds ratios for gender for Model C, controlling for risk

risk*	gender**	Odds ratio
risk = 1 (below)	gender = 1 (female)	1.65
risk = 2 (above)	gender = 1 (female)	2.14

*Note: * Reference category for risk (plan) is average-risk plan,*
*** Reference category for gender is male.*

Table A7-12: Estimated odds ratios for race for Model C, controlling for income

income*	race**	Odds ratio
income = 1 (below)	race = 1 (Asian)	1.05
	race = 2 (Black)	0.48
	race = 3 (Coloured)	0.26
income = 2 (above)	race = 1 (Asian)	n.s.
	race = 2 (Black)	n.s.
	race = 3 (Coloured)	n.s.

*Note: * Reference category for income is average income group,*
*** Reference category for race is White.*

Table A7-13: Estimated odds ratios for urban for Model C, controlling for risk

risk*	urban**	Odds ratio
risk = 1 (below)	urban = 1	1.53
risk = 2 (above)	urban = 1	2.08

*Note: * Reference category for risk is average-risk plan,*
*** Reference category for urban is rural residency*

Table A7-14: Estimated odds ratios for risk for Model C, controlling for age

risk*		age20	age40	age60	age80
OR risk	(below = 1)	1.36	1.21	1.07	0.95
	(above = 2)	0.50	0.45	0.40	0.35

*Note: * Reference category for risk (plan) is average-risk plan.*

Appendix 8

Assessment of model fit

Summary measures of goodness-of-fit. In most software packages the logistic R-squared measure or Pseudo-R-Square is available.[193] Two other goodness-of-fit tests for logistic regression models offered in most software packages are the Pearson goodness-of-fit chi-square statistic and the Hosmer-Lemeshow goodness-of-fit test.[194] As mentioned, in order to obtain the pseudo R^2, the Pearson goodness-of-fit chi-square statistic and the Hosmer-Lemeshow goodness-of-fit test, the models were run as models for uncorrelated data. On the next page table A8-1 describes the results for each models' summary measures of the goodness-of-fit which were obtained after estimating the proxy Models A, B, and C. It shows the log-likelihood, Wald test, Pearson chi-square goodness-of-fit test, Hosmer-Lemeshow goodness-of-fit test, and the Pseudo R.

For proxy Model A the value for the log-likelihood is L=-315694. The value for the Wald test is 29306 producing p-values for the test for both models which are significant at the α=0.01 level ($P(\chi^2(37)> 29306)$=0.000). The null hypothesis that at least one and perhaps all coefficients are different from zero is rejected. The log-likelihood based pseudo R values are, as expected, very low (but comparable to any other logit model). For the fit of a saturated model, R would be equal to one. Here the pseudo R for proxy Model A is pretty good with 0.12.

The Pearson chi-square goodness-of-fit test tests the observed against expected number of responses using cells defined by the covariate patterns. Since the p-values are small for both models, below 0.01, the proxy models do not seem to be fitting very well.

[193] It is a coefficient analogue to the squared contingency coefficient, with an interpretation like R-square. It's maximum is less than 1, but extremely low R^2 values in logistic regression models are the norm - typically below .1 rather than 1 - especially when compared to R^2 values encountered in good linear regression models (Hosmer and Lemeshow 2000).

[194] The Hosmer-Lemeshow goodness-of-fit chi-square statistic is calculated similarly to the Pearson chi-square goodness-of-fit statistic. The Hosmer-Lemeshow goodness-of-fit test divides subjects into deciles based on predicted probabilities and computes a chi-square from observed and expected frequencies. Then a probability (p) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model. If the test statistic is .05 or less, the null hypothesis that there is no difference between the observed and model-predicted values of the dependent is rejected. This means the model predicts values significantly different from what they ought to be, i.e. the observed values. If the test statistic is greater than .05, as desired, the null hypothesis that there is no difference cannot be rejected, implying that the model's estimates fit the data at an acceptable level (Hosmer and Lemeshow 2000). For a more detailed mathematical description of the pseudo R-squared measure and the two goodness-of-fit test statistics see Hosmer and Lemeshow (2000) and STATA Manual (1999) Vol2.

However, the number of covariate patterns is close to the number of observations, making the application of the Pearson chi-square test questionable (Hosmer and Lemeshow 2000, STATA 1999).

Table A8-1: Summary goodness-of-fit measures for proxy Models A, B, and C

Logistic proxy Model A			
Wald test	**Pseudo R2**	**Pearson test**	**Hosmer-Lemeshow test**
Number of obs = 663,644 Wald chi2(37) = 29306.07 Prob > chi2 = 0.000 Log likelihood = -315694	Pseudo R2 = 0.1453	Number of obs = 663,644 Number of covariate patterns = 663,636 Pearson chi2(663,598) = 681206.5 Prob > chi2 = 0.000	Number of obs = 663,644 Number of groups = 10 H-L chi2(8) = 181.19 Prob > chi2 = 0.000
Logistic proxy Model B			
Number of obs = 684,939 Wald chi2(37) = 17559.85 Prob > chi2 = 0.000 Log likelihood = -375510	Pseudo R2 = 0.0573	Number of obs = 684,939 Number of covariate patterns = 684,932 Pearson chi2(684,894) = 701712.5 Prob > chi2 = 0.000	Number of obs = 684,939 Number of groups = 10 H-L chi2(8) = 195.85 Prob > chi2 = 0.000
Logistic proxy Model C			
Number of obs = 532,010 Wald chi2(20) = 5340.45 Prob > chi2 = 0.0000 Log likelihood = -109592	Pseudo R2 = 0.0696	Number of obs = 532,010 Number of covariate patterns = 532,005 Pearson chi2(531,967) = 541495.6 Prob > chi2 = 0.000	Number of obs = 532,010 Number of groups = 10 H-L chi2(8) = 846.03 Prob > chi2 = 0.000

The Hosmer-Lemeshow goodness-of-fit test is more appropriate when the number of covariate patterns is close to the number of observations, as is the case in this data. This test groups the data by ordering the predicted probabilities and then forming 10 nearly equal-size groups. The value of the Hosmer-Lemeshow goodness-of-fit statistic for proxy Model A is 181. The corresponding *p*-values from the chi-square distribution with 8 degrees of freedom are all 0.000. This indicates that the proxy models are not such a good fit. However, the comparison of the observed and expected frequencies in each of the 20 cells in table A8-2 on the next page shows close agreement within each decile of risk.

Altogether, the goodness-of-fit results in table A8-1 are mixed. However, this is of no big concerns for two reasons. First, these goodness-of-fit results were obtained from the proxy models that would otherwise not be deemed adequate for panel data. Second, models are usually not entirely based on these summary measures of statistical significance.

Table A8-2: Hosmer-Lemeshow goodness-of-fit test quantiles of estimated probabilities

Group	Prob.	Obs_1	Exp_1	Obs_0	Exp_0	Total
Proxy Model A						
1	0.0576	2,815	2,379	63,550	63,986	66,365
2	0.0927	5,115	5,001	61,249	61,363	66,364
3	0.1273	7,064	7,297	59,301	59,068	66,365
4	0.1636	9,135	9,633	57,229	56,732	66,364
5	0.2041	11,802	12,167	54,562	54,197	66,364
6	0.251	14,724	15,059	51,641	51,306	66,365
7	0.3078	18,581	18,471	47,783	47,894	66,364
8	0.3839	22,776	22,806	43,589	43,559	66,365
9	0.4997	29,731	28,998	36,633	37,366	66,364
10	1	40,743	40,676	25,621	25,688	66,364
Proxy Model B						
1	0.1465	8,445	7,667	60,049	60,827	68,494
2	0.1809	11,247	11,304	57,247	57,190	68,494
3	0.2059	12,576	13,272	55,918	55,222	68,494
4	0.2287	14,609	14,890	53,885	53,604	68,494
5	0.2509	16,338	16,423	52,156	52,071	68,494
6	0.2732	17,646	17,943	50,848	50,551	68,494
7	0.2993	19,118	19,572	49,376	48,922	68,494
8	0.3377	21,770	21,711	46,724	46,783	68,494
9	0.4097	25,743	25,260	42,751	43,234	68,494
10	1	36,289	35,740	32,204	32,754	68,493
Proxy Model C						
1	0.0115	451	492	52,750	52,710	53,201
2	0.0203	963	916	52,238	52,285	53,201
3	0.0239	855	1,175	52,346	52,027	53,201
4	0.0348	2,062	1,509	51,139	51,692	53,201
5	0.0558	2,936	2,304	50,265	50,897	53,201
6	0.0716	3,073	3,454	50,128	49,747	53,201
7	0.0808	3,223	4,068	49,978	49,133	53,201
8	0.0897	4,074	4,527	49,127	48,674	53,201
9	0.1028	5,204	5,079	47,997	48,122	53,201
10	1	8,011	7,329	45,190	45,872	53,201

Often it is desirable to examine individual components of the summary statistic, for example with classification tables. Classification tables tally correct and incorrect estimates. They cross-classify the outcome variable with a dichotomous variable whose values are derived from the estimated logistic probabilities. A cut-point is defined (most commonly 0.5) in order to compare each estimated probability to it. Table A8-3 shows the results for classifying the observations with proxy Models A, B, and C. The common cut-off point 0.5 was used here. The columns of the table contain the two predicted values of the dependent variable, while the rows are the two observed (actual) values of the dependent variable. In a perfect model, all cases would be on the diagonal and the overall percent correct will be 100%. If the logistic model has homoscedasticity, the percent

correct will be approximately the same for both rows. Sensitivity is the fraction of the outcome of y_i=1 observations that are correctly classified. Specificity is the percent of y_i=0 observations that are correctly classified.

For proxy Model A the overall rate of correct classification is estimated as 78%. The specificity is 95% (475,599/501,158), meaning 95% of the joiner group being correctly classified. The sensitivity is 25% (40,672/162,486), meaning only 25% of the control group being correctly classified. For proxy Model B the overall correct classification rate is 74% and for proxy Model B 94%. Since the specificity in proxy Models B and C are even higher at the 0.5 cut-point, the sensitivity must be lower.

Table A8-3: Classification table and classification statistics for proxy Models A, B, and C

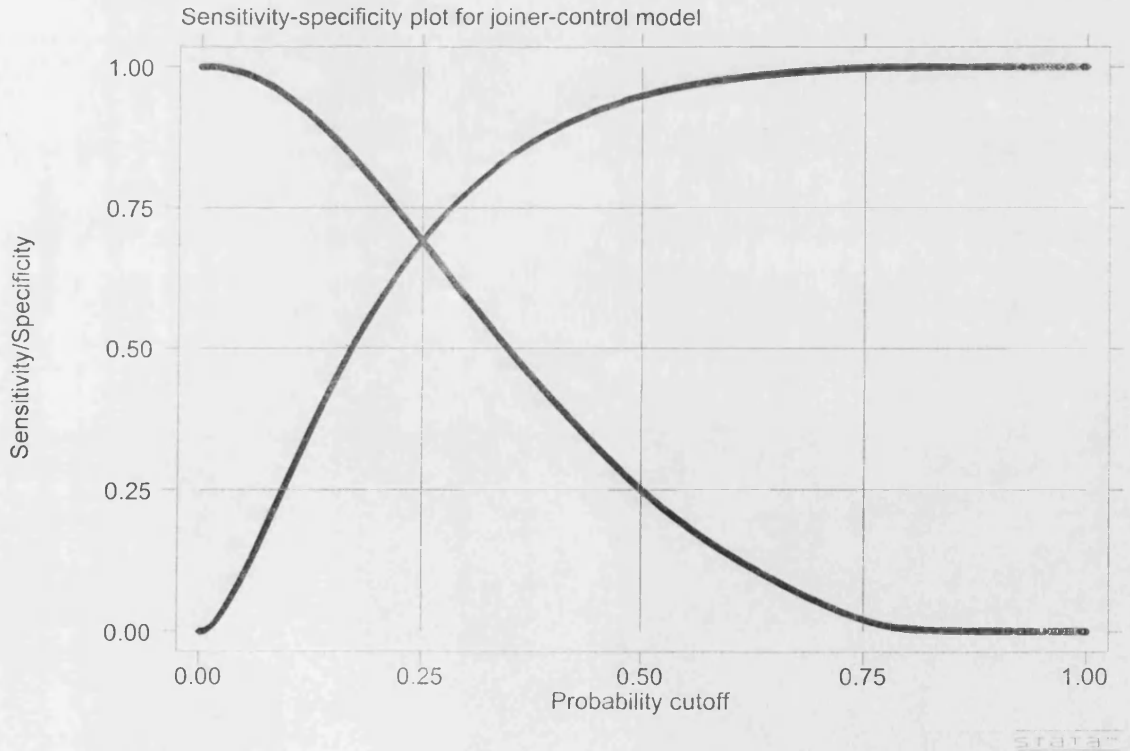
	Proxy Model A			Proxy Model B			Proxy Model C		
	True			True			True		
Classified	D	~D	Total	D	~D	Total	D	~D	Total
+	40,672	25,559	66,231	19,305	12,123	31,428	6	28	34
-	121,814	475,599	597,413	164,476	489,035	653,511	30,846	501,130	531,976
Total	162,486	501,158	663,644	183,781	501,158	684,939	30,852	501,158	532,010
Classified + if predicted Pr(D) >= .5									
True D defined	jc ~= 0			lc ~= 0			sc ~= 0		
Sensitivity	Pr(+| D)	25.03%		Pr(+| D)	10.50%		Pr(+| D)	0.02%	
Specificity	Pr(-|~D)	94.90%		Pr(-|~D)	97.58%		Pr(-|~D)	99.99%	
Positive predictive value	Pr(D| +)	61.41%		Pr(D| +)	61.43%		Pr(D| +)	17.65%	
Negative predictive value	Pr(~D| -)	79.61%		Pr(~D| -)	74.83%		Pr(~D| -)	94.20%	
False + rate for true ~D	Pr(+|~D)	5.10%		Pr(+|~D)	2.42%		Pr(+|~D)	0.01%	
False - rate for true D	Pr(-| D)	74.97%		Pr(-| D)	89.50%		Pr(-| D)	99.98%	
False + rate for classified +	Pr(~D| +)	38.59%		Pr(~D| +)	38.57%		Pr(~D| +)	82.35%	
False - rate for classified -	Pr(D| -)	20.39%		Pr(D| -)	25.17%		Pr(D| -)	5.80%	
Correctly classified		77.79%			74.22%			94.20%	

Table A8-3 also demonstrates that for proxy Model A 475,599 of the 597,413 households predicted to be in the control group were classified correctly, while 25,559 of the 66,231 households predicted to be joiners were misclassified. Of the total number of 162,486 households who were actually joiners, only 40,672 were correctly predicted. This considerable amount of misclassification suggests that many households in the data have probabilities close to the cut-point (Hosmer and Lemeshow 2000). If these proxy models

would have been actually fit to cross-sectional data (and not this thesis' panel data) they would point towards a high level of homogeneity in the described insured population.

However, classification table results generally do not have a lot of interpretation value for the goodness-of-fit of a model, because they do not assess the calibration, i.e. if probabilities reflect the true outcome experience in the data (Hosmer and Lemeshow 2000). But often it is the objective to choose an optimal cut-point for the purposes of classification. This optimal cut-point is where sensitivity and specificity are both maximised. First, one can chose a range of different possible probability cut-off points for the previously discussed classification tables, then obtain their respective sensitivity and specificity values and finally plot sensitivity and specificity versus all possible cut-points, as in figure A8-1. The intersection of both curves then displays the optimal cut-point for the purpose of classification. In the plot under figure A8-1 sensitivity and specificity were plotted against all possible probability cut-off points of Model A. For proxy Model A the plot indicates that the optimal cut-off point would be 0.25.[195]

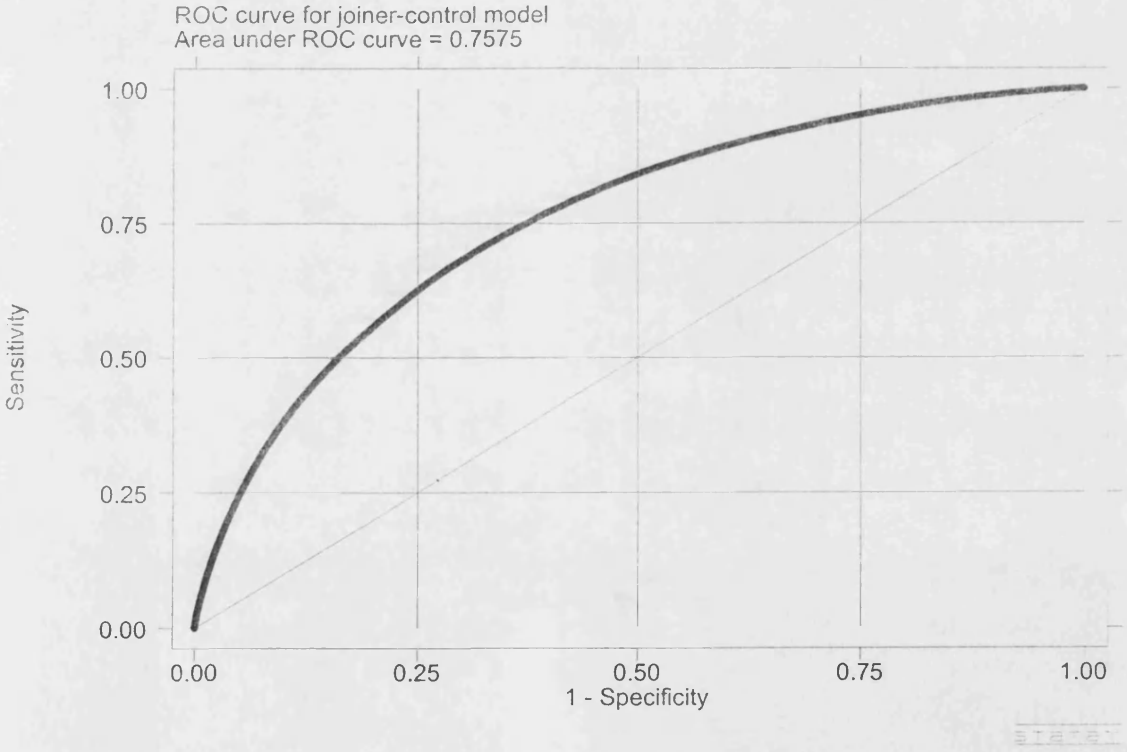
Figure A8-1: Plot of sensitivity and specificity versus all possible cut-points in proxy Model A



[195] For proxy Models B and C the optimal cut-off point would be 0.26 and 0.075 respectively.

A more comprehensive description of the classification accuracy than the single cut-point, which sensitivity and specificity rely upon, is the area under the so-called ROC (Receiver Operating Characteristics) curve. The ROC curve is generated by plotting the sensitivity (or fraction of observed positive-outcome cases that are correctly classified) versus 1-specificity (or fraction of observed negative-outcome cases that are correctly classified) over the entire range of possible cut-points. Values under the curve .6 ≤ ROC ≤ .9 are most common, indicating acceptable to excellent discrimination. Proxy Model A has an area under the ROC curve of 0.76, proxy Model B 0.66, and proxy Model C 0.71, respectively. This area under the curve provides as a measure of discrimination: the likelihood that a household who joins/leaves/switchers will have a higher probability y=1 than a household who is in the control group. The plot in figure A8-2 displays the ROC curve for proxy Model A with areas under the ROC curve that indicate a good discrimination.

Figure A8-2: ROC–curve for proxy Model A



Logistic regression diagnostics. If the independent variables of two observations are identical these two observations share the same covariate pattern. Then, despite having individual observations, the statistical information of the data can be summarised by the covariate patterns (i.e., the number of observations with that covariate pattern and the

number of positive outcomes within the pattern) (STATA 1999).[196] Proxy Model A, for example, was estimated over n=663,644 household observations with J=663,636 covariate patterns.[197]

STATA™ calculates all residuals and diagnostics statistics in terms of covariance patterns, not observations. Observations with the same covariate pattern are given the same residual and diagnostic statistics. The Pearson chi-square statistic, introduced under the summary measures of goodness-of-fit, can be then computed based on all observations with this covariate pattern.[198] As an example, table A8-4 describes the Pearson residuals for proxy Model A based on all observations with a specific covariate pattern. Prevalence of a few, large positive residuals is noticeable.

Table A8-4: Pearson residual based on observations with similar covariate pattern for proxy Model A

Proxy Model A	Pearson residual			
	Percentiles	**Smallest**		
1%	-1.350	-71.514		
5%	-0.938	-24.741		
10%	-0.774	-14.638	**Obs.**	663,643
25%	-0.560	-14.174	**Sum of Wgt.**	663,643
50%	-0.367		**Mean**	0.001
		Largest	**Std. Dev.**	1.013
75%	-0.110	14.421		
90%	1.532	14.699	**Variance**	1.026
95%	2.033	18.606	**Skewness**	1.267
99%	3.275	20.074	**Kurtosis**	46.169

Pearson residuals do not have standard deviation equal to one unless they are further standardised (Hosmer and Lemeshow 2000, STATA 1999). However, generated Pearson residuals, normalised to have expected standard deviation equal to one, do not differ for this thesis' data, as is demonstrated in table A8-5. Thus the question of whether to use standardised or unstandardised residuals does not matter here.

[196] It can be assumed that a fitted model contains p covariates that form J covariate patterns. Depending on the model the number of covariate patterns can be less or be equal to n, the number of observations. The fitted models in this thesis contained several continuous covariates; hence the number of covariate patterns J is of the same order as n.

[197] Models B was estimated over n=684,939 and J=684,932 covariate patterns, Model C had n=532,010 and J=532,005.

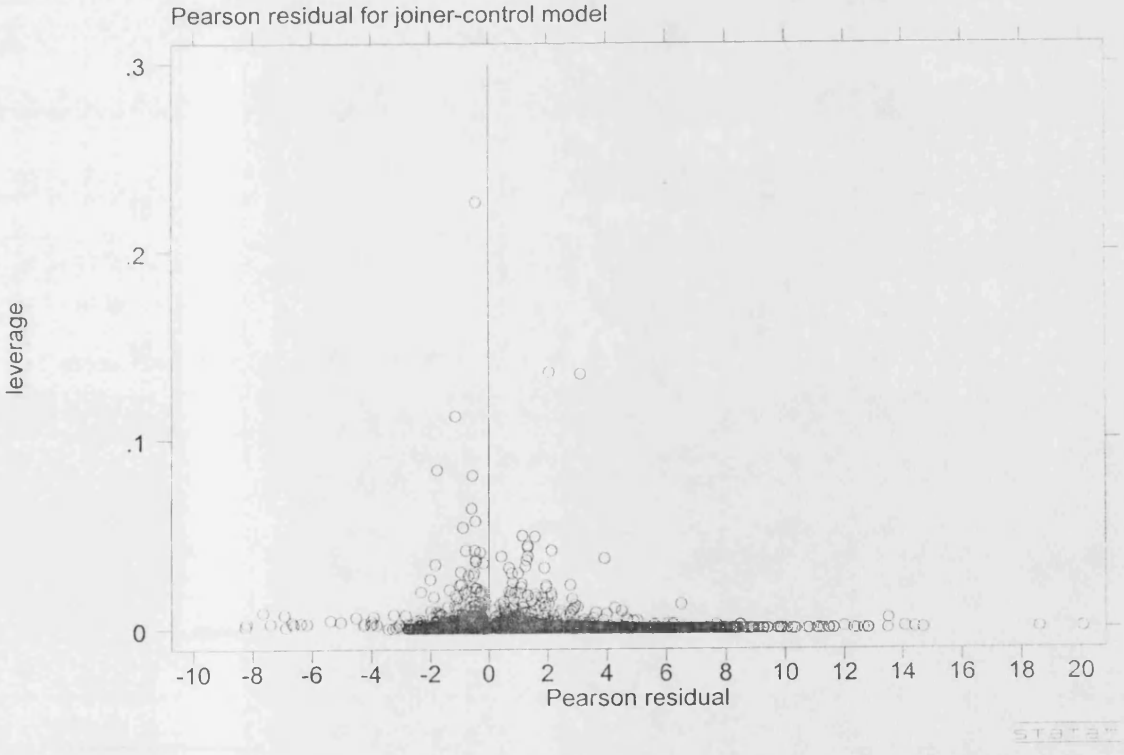
[198] The residual is defined as the square root of the contribution of the covariate pattern to the Pearson chi-square goodness-of-fit statistic and signed according to whether the observed number of positive responses within the covariate pattern is less or greater than expected.

Table A8-5: Summary of unstandardised and standardised residuals for proxy Model A

Variable	Obs.	Mean	Std. Dev.	Min	Max
r – unstandardised	663,643	0.001	1.013	-71.514	20.074
rs – standardised	663,643	0.001	1.013	-71.516	20.076

Figure A8-3 presents the plot of the leverage values versus the Pearson residuals for proxy Model A.[199] The leverage values are the diagonal elements of the hat matrix and this calculated leverage of a covariate pattern is a scaled measure of distance in terms of the independent variables.[200] Large values indicate covariate patterns deviating strongly from the average covariate pattern that can have a large effect on the fit of an estimated model (STATA 1999). This thesis' data contains almost as many covariate patterns as observations, thus displaying many unique covariate patterns (in figure A8-3). In such unique patterns one observes either zero or one success and expects p, thus forcing the sign of the residuals (STATA 1999). Covariate patterns that are not fitted well by the model are represented by large residuals and appear on the left and right edges of figure A8-3.

Figure A8-3: Plot of leverage versus the Pearson residual for proxy Model A

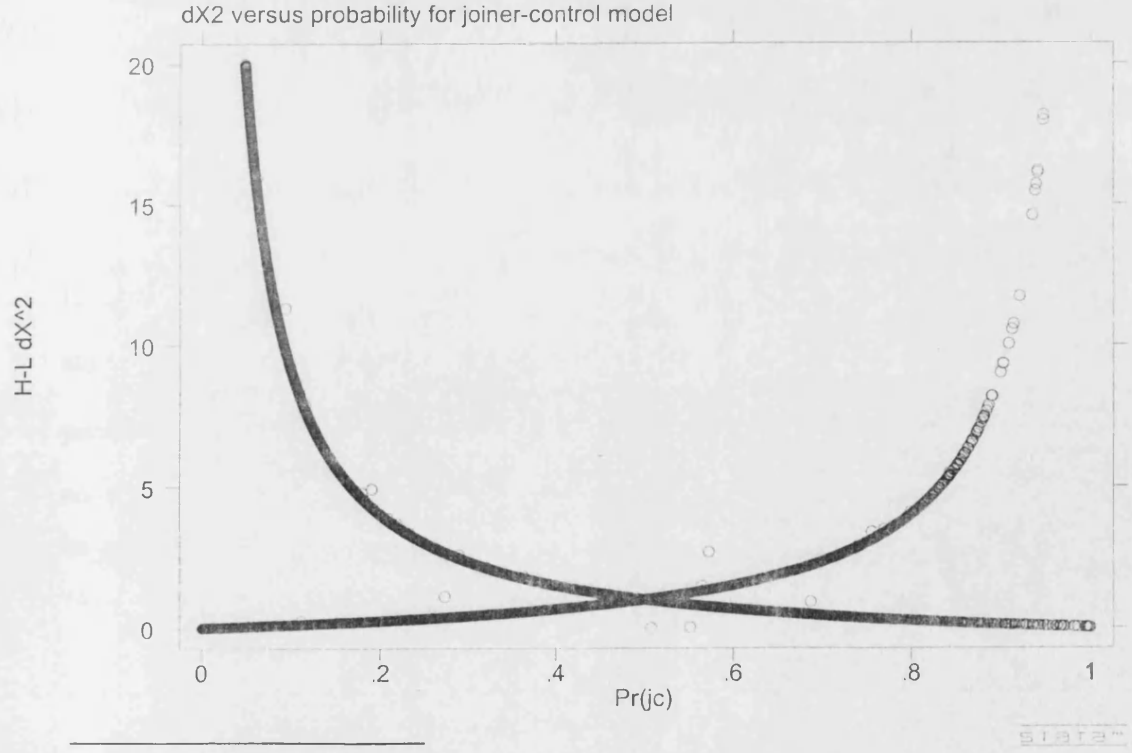


[199] The plots for proxy Models B and C were similar and are not displayed here.
[200] For a more detailed mathematical development see Hosmer and Lemeshow (2000) p 167ff.

Covariate patterns that appear on the top of figure A8-3 have a high leverage. Points on the extreme right and left represent large residuals. Covariate patterns with high leverage and small residuals are of most interest because they are hard to spot with the previous diagnostics. Overall, however, figure A8-3 shows only a very small number of deviating covariate patterns, indicating a good model fit.

Figure A8-4 presents the plot of the Pearson chi-square goodness-of-fit statistic $\Delta\chi^2$ against the estimated logistic probability residuals for proxy Model A.[201] The shape of the plots show quadratic curves, the points on the curve from top left to bottom right correspond to covariate patterns with y_j=1 and the points on the other curve from the bottom left to top right correspond to covariate patterns with y_j=0.[202] Points that fall into some distance from the data plot, on the top left or top right, represent some covariate patterns that are poorly fit. However, they are generally few poorly fit covariate patterns and most values fall under or around 4. This means most values of $\Delta\chi^2$ fall under or around 4 (these quantities would be distributed approximately as chi-square with one degree of freedom with $\chi_{2.95}(1) = 3.84$) and thus the plot indicates a reasonable model fit.

Figure A8-4: Plot of $\Delta\chi^2$ versus the estimated logistic probability from proxy Model A

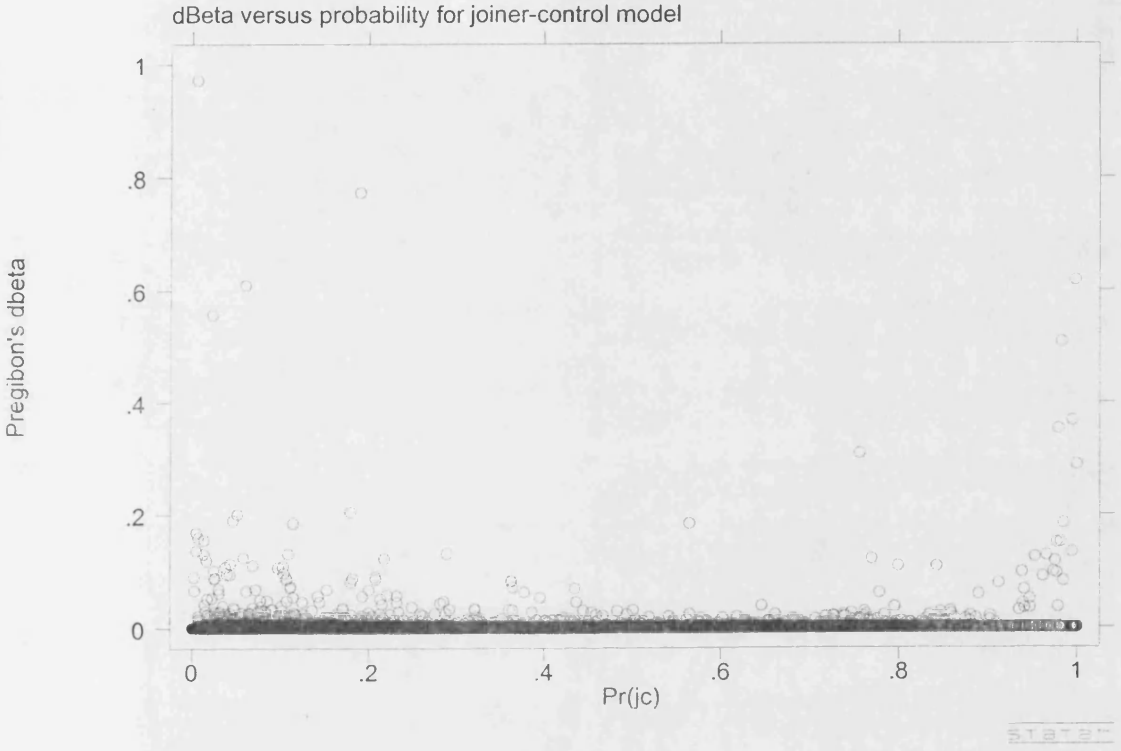


[201] The plots for proxy Models B and C were similar and are not displayed here.

[202] In the data most covariate patterns are unique and points tend to lie along one or the other curve while points off the curves correspond to the few repeated covariate patterns that contain different outcomes.

In figure A8-5 the influence diagnostic $\Delta\hat{\beta}$ is plotted against the estimated logistic probability for the proxy Model A.[203] In general the influence of diagnostics $\Delta\hat{\beta}$ has to be larger than one to affect the estimated coefficients (Hosmer and Lemeshow 2000). There are only a few points that lie away from the data, indicating a good model fit.

Figure A8-5: Plot of $\Delta\hat{\beta}$ versus the estimated logistic probability from proxy Model A



Large values of $\Delta\hat{\beta}$ most likely occur when both $\Delta\chi^2$ and leverage, or either of the two, are comparably large. This can be assessed by plotting $\Delta\chi^2$ against the estimated logistic probability with the size of the symbol proportional to $\Delta\hat{\beta}$, shown in figure A8-6 on the next page (for proxy Model A). A larger circle on the upper left indicates large $\Delta\chi^2$ values, larger circles for the estimated probabilities around 0.2 indicate the region of high leverage values. Overall, however, figure A8-6 indicates a very good model fit.[204]

[203] The plots for proxy Models B and C are similar and are not displayed here.
[204] Similar plots were generated for proxy Models B and C, however, they are not displayed here.

Figure A8-6: Plot of $\Delta\chi^2$ versus the estimated logistic probability (symbol proportional to $\Delta\hat{\beta}$) from proxy Model A