Predicting the Expected Costs of Health Care
Methods for Risk Adjustment in Health Services

Full Report

Dr Stuart Peacock
Senior Lecturer,
Health Economics Unit, Centre for Health Program Evaluation

Dr Leonie Segal
Deputy Director
Health Economics Unit, Centre for Health Program Evaluation

Professor Jeff Richardson
Director
Health Economics Unit, Centre for Health Program Evaluation
The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding in the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of General Practice and Public Health at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

---

**PUBLICATIONS**

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator  
Centre for Health Program Evaluation  
PO Box 477  
West Heidelberg  Vic  3081, Australia  
**Telephone** + 61 3 9496 4433/4434  
**Facsimile** + 61 3 9496 4424  
**E-mail** CHPE@BusEco.monash.edu.au  
ACKNOWLEDGMENTS

This project was funded by the Commonwealth Department of Health and Aged Care and undertaken by the Health Economics Unit, Monash University. The views expressed in the Report are those of the authors and are not necessarily the views of the Commonwealth Department of Health and Aged Care.

The Health Economics Unit of the CHPE is supported by Monash University.

The Program Evaluation Unit of the CHPE is supported by The University of Melbourne.

Both units obtain supplementary funding through national competitive grants and contract research.

The research described in this paper is made possible through the support of these bodies.

AUTHOR ACKNOWLEDGMENTS

The authors would like to thank Mr Peter Broadhead and Ms Suzie Northcott, Commonwealth Department of Health and Aged Care, and Mr Ron Donato, Adelaide University for their input and comments at various stages of the project.
# Table of Contents

Executive Summary .................................................................................................................. i
A Key Observations.................................................................................................................. i
B Research Questions for Australia........................................................................................ iv

1 Introduction ............................................................................................................................. 1

2 Predicting Health Service Costs ............................................................................................. 2
  2.1 Risk Adjustment Methods for Predicting the Costs of Health Care ................................ 2
  2.2 Criteria for Assessing Risk Adjustment Methods ......................................................... 11
  2.3 Normative Approaches to Risk Adjustment ................................................................. 16
  2.4 Historic Expenditure Approaches to Risk Adjustment ................................................ 17

3 Individual Level Risk Adjustment Methods .......................................................................... 19
  3.1 Risk Adjustment Using Individual Level Analysis ......................................................... 19
  3.2 Validity and Predictive Power ......................................................................................... 23
  3.3 Validity of Needs Factors ............................................................................................... 25

4 Aggregate Level Risk Adjustment Methods ........................................................................ 30
  4.1 Risk Adjustment Using Aggregate Level Analysis ......................................................... 31
  4.2 Validity of Needs Factors ............................................................................................... 36

5 Generic Issues with Risk Adjustment Methods .................................................................. 37
  5.1 Unmet Need ....................................................................................................................... 37
  5.2 Normative Utilisation ......................................................................................................... 38
  5.3 Public-Private Mix ............................................................................................................. 39
  5.4 Urban and Rural Populations ............................................................................................ 40
  5.5 Cross-Boundary Flows ....................................................................................................... 42

6 Supply Side Factors and Risk Adjustment ........................................................................... 43
  6.1 Competitive Health Scheme Models ................................................................................ 43
  6.2 Non-Competitive Health Scheme Models ....................................................................... 44
  6.3 A Synthesised Approach to Capitation ........................................................................... 45
  6.4 Supply Side Factors in Empirical Studies ....................................................................... 48

7 Risk Adjustment for Capitation Funding ............................................................................. 50
  7.1 Budget Setting and Scheme Coverage ............................................................................. 50
  7.2 Financial Risk and Risk Pooling ....................................................................................... 52
  7.3 Ex Post Risk Sharing Arrangements ................................................................................ 55
  7.4 Implementation Strategies for Risk Management ........................................................... 57

8 Overview of Capitation Issues in Australia ......................................................................... 58
  8.1 Capitation Funding and Australian Health Services ....................................................... 58
  8.2 Risk Adjustment Methods and Australian Health Services .......................................... 59
  8.3 Risk Adjustment and the Australian Coordinated Care Trials ...................................... 65
  8.4 Research Questions for Australia .................................................................................... 67

References .................................................................................................................................. 70
List of Tables

Table 2.1  Examples of needs factors................................................................. 6
Table 2.2  Examples of supply side factors......................................................... 7
Table 2.3  Capitation experiences in 19 countries (from Rice and Smith 1999)......... 9
Table 2.4  Potential responses of health schemes to incentives under capitation....... 13
Table 2.5  Potential distortions in health scheme data reporting............................ 15
Table 3.1  Comparison of $R^2$ Values................................................................. 29
Table 4.1  York Acute and Psychiatric Models...................................................... 35

List of Figures

Figure 4.1  The York Model of the Demand for Health Care................................. 32
Figure 8.1  Funds Pool model, typical models and the Australian CCTs.................... 67
Executive Summary

A Key Observations

(i) Risk Adjustment Methods for Predicting the Costs of Health Care  
(Sections 2-6)

There is a substantial world-wide research agenda to develop risk adjustment methods to predict the ‘expected’ costs of health services. The methods are used to develop risk-adjusted capitation funding models for health services. The modelling of health service utilisation and cost therefore represents a major research activity in many countries. The primary purpose of this research is the development of risk adjusted capitation-based funding formulae, to determine the distribution and level of health care resources. Capitation funding formulae determine prospective budgets for ‘health schemes’ based on the ‘expected’ costs of health service use of enrollees in those schemes. ‘Health schemes’ are agencies that are responsible for the delivery of health services for a defined population at risk over a given time period, and may be public or private sector bodies. The methods used to estimate the expected costs of health service use are collectively known as risk adjustment methods. Risk adjustment is the use of information to calculate the expected health expenditures of individuals over a fixed period of time to determine payment rates for health schemes.

Risk adjusted capitation formulae may encompass just acute care services, primary care services, or all health services, (variously defined). Risk adjustment methods and capitation funding are used in both non-competitive health scheme models (for example Canada, the United Kingdom and Australia) and competitive health scheme models (for example the United States and the Netherlands). Competitive models allow health schemes to compete for enrollees (for example under private health insurance or managed competition), whereas under non-competitive models competition for enrollees is not permitted (for example where health schemes have a defined geographical catchment area).

Risk adjustment methods have concentrated on the development of regression models to explain variations in health service utilisation and cost in terms of the demographic, health, and social characteristics of individuals (needs factors) and the organisational, managerial and policy characteristics of health service providers (supply factors). These models have been used to predict the expected costs of health care for individuals, accounting for interpersonal differences in the need for health services, and in some instances, differences in the supply side characteristics of health service providers.

A rigorous approach to the research task is considered critical, to engender confidence in the resulting formulae in terms of its robustness, transparency, and ability to provide appropriate incentives.

---

1 As discussed in this Paper `expected' expenditure is variously interpreted. While the simple interpretation is the mean health service use and cost implied by a particular mix of patient characteristics - the extent to which provider characteristics and behaviour (and costs), and broader health system delivery and funding arrangements should be brought into consideration vary.
Inappropriate risk adjustment methods can result in insolvency for health schemes, or, in the competitive context, perverse incentives for risk selection (cream skimming and adverse selection). Cream skimming and adverse selection refer to the practices of attracting enrollees for whom the risk adjusted payment is greater than expected cost, and to discouraging enrollees where payment is less than expected cost.

(ii) Individual Level versus Aggregate Level Analysis (Sections 3 & 4)

Risk adjustment models have been based either on individual/patient level data or small area data (such as census collector district or postcode). Typically small area data has been used in the non-competitive health scheme models. This is primarily because individual level data is less commonly found in health systems where individual insurance claims data is not required. However, under a non-competitive model, it is also less critical to establish a correct capitation weight for each individual, because there is limited potential for risk selection where enrollment is dictated by factors outside health scheme control (such as geographic location), and budgets are typically set for a defined population group.

In theory, the use of individual level data is preferable, regardless of health scheme type. Individual level analysis allows for more precise estimation of individual’s expected costs, because it allows more accurate description of the relationship between health service use and costs, and individuals’ demographic, health and social characteristics. Small area data does not capture the full extent of heterogeneity between individual’s within small areas, and may result in less precise predictions of costs at the individual level.

(iii) Capacity of Models to Predict Health Service Costs (Sections 3, 4, 6 & 7)

The capacity of models to predict health service use and cost depends primarily on the structure of the model, the choice of explanatory variables, whether individual or small area data is used, and the scope of services covered. Whether the capitation formula is consistent with the financial survival of health schemes will depend on the extent to which unexplained variation is a reflection of random error (rather than statistical bias), scheme size in terms of number of enrollees, and strategies for risk sharing.

Using individual level data, it has been estimated that capitation models can predict up to 20-25% of variation in health service costs. This is widely held to represent the proportion of health service costs that is potentially explicable under statistical analysis. The remaining 75-80% of variation in costs is due to random factors. In practice, many capitation formulae based on individual level data have reported that 10-12% of the variation in costs was explained in regression modelling. Studies of selected types of health services have demonstrated less random variability in costs for some services, for example chronic care and pharmaceuticals, and have reported as much as 56% of variation in costs explained in modelling.

In relation to financial survival of health schemes, even risk adjustment models which explain only 10% of individual level variation will be likely to cover actual costs, provided the enrolled population is of sufficient size (and budgets reflect predicted cost).
Whilst little empirical work has been carried out on optimal scheme size for capitation formula based on individual level data, an enrolled population of 5,000 is suggested as sufficient to minimise the risk of financial insolvency due to random influences on costs.

The percent of variation in health service cost predicted by models using small area data tends to be higher, at around 55%. This is not because of any superiority in the model structure, but because the combination of data, on the cost side as well as in relation to independent variables, removes substantial variation in the data. However, in translating to financial robustness of schemes, far larger enrolled populations are required. At an enrolled population of 10,000 persons it has been estimated that approximately 17% of Schemes would find expenses more than 10% in excess of capitation payments, (and when small areas used in analysis have an average population of 10,000). This proportion falls rapidly to 0.1% for an enrolled population of 100,000 persons.

(iv) Typical Explanatory Variables in Risk Adjustment Methods (Sections 2-4)

Variables used in capitation models are drawn from a range of measures for needs factors and supply side factors.

Typical needs factor variables include:

Demography Age, Sex, Ethnicity
Employment/ Employment, Unemployed, Pensioner, Disability Status Temporarily unable to work, Permanently sick.
Health Status Self reported/survey-based morbidity, Mortality rates, Permanent disability/dependency status, Low birth weight Previous inpatient/outpatient diagnosis, (DCG diagnostic cost group, ACG ambulatory care group etc).
Socioeconomic Status Homelessness, Martial status, Income, Housing tenure Socioeconomic Status/Social class, Education level Elderly living alone, No carer in household.
Geographic location Region/area of residence, Rurality, Urbanisation Remoteness, Population Density

Typical supply side factor variables include:

Provider Characteristics Input prices, Labour prices, Capital rental prices (Exogenous to health scheme) Location specific service delivery costs Rurality, Urbanisation, Remoteness
Provider Characteristics Market power - ability to get price discounts (Endogenous to health scheme) Insurance coverage - deductibles, co-payments etc Contracting arrangements, Practice style Health scheme management policies and practices
(v) Approaches to Risk Sharing (Section 7)

There are several possible approaches to risk sharing. These are necessary because health service costs are subject to significant random variation, and risk adjustment methods may only imperfectly estimate the relationship between needs and supply factors, and health service costs. Both may expose health schemes to significant financial risk. Increased scheme size is probably the most important means to contain risk. But even in larger schemes a small risk of insolvency may remain. Risk may also be reduced through exclusion from the capitation formula and scheme responsibility of nominated high cost activities, and/or the most expensive x% of cases. The presumption is that the costs of such ‘outliers’ will be met at the funder/national level (or at least beyond the scheme level).

(vi) Population and Service Coverage (Section 7)

Little attention has been given to the scope of services to be included in health schemes. However, this has significant implications for the development of risk adjusted formulae and incentives for cost shifting. In general, the narrower the scope of services covered, the greater the capacity for cost-shifting to services not included in the formula. But where quite distinct service types are to be covered, (such as acute care, mental health, primary care, residential care), it is probable that a number of distinct capitation formula will need to be developed.

B Research Questions for Australia

A major function of this Paper is to stimulate and contribute to an on-going dialogue about health system funding and delivery, and the role for risk adjustment and capitation funding in Australia. The types of issues that will ideally be explored as part of this debate are listed below:

(i) Health System Objectives and the Role of Capitation Funding

A robust debate about the objectives of the health system and the possible contribution of risk adjusted capitation to those objectives is desirable. It is of particular importance to ascertain the views of policy makers and the community about the relative importance of efficiency, equity, and access. While sometimes equity and efficiency objectives will be complementary, there may also be conflict between goals, requiring choices and trade-offs. Such trade-offs require value judgements, which cannot be determined through regression based risk adjustment modelling. This matter is central to the development of an appropriate capitation rate. It is necessary to determine what it is designed to do. Is it merely to replicate past expenditures and historic access to services, or is it to be used to redistribute resources to ensure a fairer access to services - however that is defined?

(ii) Supply Side Factors and Risk Adjustment

Alternative stances to the treatment of avoidable cost differences (at the health scheme and provider level) have been adopted in the risk adjustment methods applied. Avoidable cost differences refer to these supply side factors endogenous to the health scheme – within the scheme’s control – including managerial efficiency and local policies and practices. Potential sources of avoidable cost differences include, historical inequities in access, and variations in
prescribing patterns and procedure rates which cannot be justified in terms of need, cost-effectiveness or equity grounds. Essentially risk adjustment may seek to include or exclude avoidable cost differences from capitation funding.

Those health systems that have included avoidable cost differences in capitation payments have done so primarily to reduce the potential for risk selection by health schemes. Attempts to correct for avoidable cost differences under capitation may promote cream skimming behaviour. If a relatively inefficient (high cost) provider is reimbursed at the average efficiency rate of payment then it will seek to reduce costs (improve efficiency). One potential means of reducing costs is to cream skim. Therefore, attempts to influence supply side behaviour through capitation payments may create perverse incentives that run contrary to the underlying goal of capitation in a competitive health scheme model. This has led to a focus on capitation payments that reflect variation in costs in competitive health scheme models, irrespective of the source - need or supply side - of that variation.

Those health systems (mainly non-competitive) that have excluded avoidable cost differences have done so primarily in the pursuit of equity objectives. Capitation to promote equity has focussed on payment rates based on predicting costs in terms of individuals’ health service needs, where rates are corrected, or adjusted, for avoidable cost differences. The correction for avoidable costs implies each health scheme receives payment based on an individual’s health service needs given some standard level of circumstances. The reimbursement rate reflects the expected cost for an individual with a given level of need adjusted for the average impact of variations in clinical policy, managerial efficiency, and the availability of health services on costs. For example, for a given level of population need, a health scheme that has higher utilisation (and hence costs) because of historically high access to services receives the same payment as a scheme with low utilisation due to historically poor access. In this way, capitation may be used to correct existing inequities in service provision between health schemes, by using relative resource shifts from high access to low access areas.

Each approach may be valid for the stated objectives of capitation funding in alternative contexts. A discussion is therefore required in the Australian context about the objectives of the health system and any potential developments in capitation funding. A number of options exist that may represent valid objectives and approaches to avoidable cost differences. The primary objective of risk adjusted capitation may be to pool funds, in the pursuit of gains in efficiency from service substitution. Equally, risk adjusted capitation may seek to reduce any potential for risk selection and quality scrimping. In both cases it may be appropriate to include avoidable cost differences in capitation payments. However, capitation funding may also be used to promote equity in access to health services. In this case it may be appropriate to exclude avoidable cost differences.

Experiences from overseas suggests that these, and other, potential objectives require considerable debate and research before a robust capitation model can be implemented. Research issues include the extent to which these objectives are relevant to the Australian context, the degree to which objectives are complementary, what trade-offs between objectives may be required, and the development of risk adjustment methods appropriate for any given set of objectives.

(iii) Population and Service Coverage
Decisions need to be made about the services to be provided by health schemes and therefore included in risk adjusted capitation formulae. Options include for instance, primary medical care only, all primary care, primary plus secondary, plus selected or all tertiary care, plus residential care. It may also include selected components of disability and welfare services. These decisions should reflect the potential for improvements in allocative efficiency through service substitution from the removal of budget boundaries.

(iv) Capped or Open-Ended Budget/Treatment of Private Health Insurance

In adopting a risk adjusted capitation based model of funding there is a presumption that health service costs can be constrained by the scheme management. Costs which are outside the control of scheme management, might be determined outside the capitation arrangement. This might include both regional based supply side variables such as input prices, or travel costs associated with distance, but might also need to accommodate national commitment to unrestricted access to certain types of services. Thus access to private medical services included on the MBS and PBS are guaranteed, and not able to be the subject of restriction. Similarly conditions of access to public hospitals are dictated by joint Commonwealth State Agreements.

Access to services funded through private health insurance may also not be within the control of scheme managers. How these issues can be, and are to be, reconciled within a risk adjusted capitation funding model need to be explored.

(v) Support for a Research Program to Develop Capitation Formulae

The wider adoption of capitation based funding is an important possible future direction for the Australian health system. Progress in this area relies heavily on two types of research:

- into the exploration and development of risk adjusted capitation funding models
- into the characteristics of the Australian health system, and how or whether they can be reconciled with risk adjusted capitation funding

The three countries which have led the way in developing risk adjustment methods, and in successfully implementing capitation funding models – the US, the Netherlands and the UK – have all devoted significant time and resources to these key research tasks. Experiences have been remarkably similar in all three countries: that a core group of three to five researchers working over a period of two to three years is required, as a minimum, to develop methods and analyses to construct robust risk adjusted capitation formulae, that will provide appropriate incentives consistent with health system objectives.
1 Introduction

A key feature of the development of publicly and privately financed health systems has been the emergence of health ‘schemes’. Despite major differences between health systems, these schemes possess a common theme: they are responsible for the delivery of health services for a defined population at risk over a given time period. They are largely the result of a global trend of devolving responsibility for health service budgets and delivery away from central government, which has occurred irrespective of the means of raising public finance, for example social insurance or taxation.

Health schemes have developed in many different forms, including insurance pools (US), sickness funds (Netherlands, Germany), and geographic areas (UK, Sweden). In order to provide services, health schemes receive a budget allocated from funders for which they are responsible. Often the funder is the central government, but in some cases other agencies, for instance state or local government or private insurers, may fund all or part of health service delivery. An integral part of the budget holding responsibility, is that health schemes bear some or all of the financial risk associated with the random variation in health expenditures across its enrollees. The devolution of budgets and service delivery responsibilities has led to increased scrutiny of the behaviour of health schemes with respect to expenditure control. Increasingly, emphasis has been placed on the determination of prospective budgets to allocate funds on a rational basis to health schemes, from which their populations’ health services are to be provided.

The determination of prospective budgets requires estimation of the expected costs of health service use for the population covered by each health scheme budget. These predicted costs are used to set capitation rates – dollar amounts for each individual enrolled in a scheme – which typically indicate how much a health scheme would spend on an individual if it were to provide some standard level of health services. The prospective budget for each scheme is therefore made up of the number of enrollees weighted for their expected costs of health service use. The methods used to estimate the expected costs of health service use are collectively known as risk adjustment methods. Risk adjustment is the use of information to calculate the expected health expenditures of individuals over a fixed period of time to determine payment rates for health schemes. This paper seeks to synthesise the literature on risk adjustment methods and the prediction of the expected costs of health care, and to provide a discussion of the relative merits of alternative approaches. Issues of particular relevance to the Australian context are highlighted.

The paper is organised as follows. The underlying principles of risk adjustment and predicting health service costs are discussed in section 2, including discussion of the criteria over which alternative approaches to risk adjustment should be assessed and an examination of historic expenditure and normative approaches to risk adjustment. Sections 3 and 4 provide more detailed discussion of individual and aggregate level risk adjustment methods, focussing on key advances in the development of empirical approaches to risk adjustment and predicting health service costs. In Section 5 is then discussed a range of generic methodological issues common to all empirical approaches to risk adjustment. Section 6 examines the treatment of supply side factors in risk adjustment and predicting health service costs.
In Section 7 is discussed key issues in the application of capitation rates in risk adjusted capitation funding models - including service and population coverage, risk pooling, and risk sharing strategies. Finally, in Section 8 much of the preceding material is brought together to examine the issues raised in relation to the Australian health services context and the Australian Coordinated Care Trials.

2 Predicting Health Service Costs

In recent years funding agencies, whether public or private sector bodies, have paid increasing attention to expenditure control in the delivery of health services. In particular, funding agencies have focussed on appropriate mechanisms to secure expenditure control within budget holding health schemes. This has led to a growing emphasis on the determination of prospective budgets to allocate funds on a rational basis to health schemes, from which their populations’ health services are to be provided. The focus of this paper is on the methods used to predict health service costs to determine prospective funding levels.

2.1 Risk Adjustment Methods for Predicting the Costs of Health Care

A number of methods may be used to estimate expected health service costs and determine prospective funding levels, including bilateral negotiations between the funder(s) and health schemes, or extrapolation of historical expenditure levels. Both of these approaches have been heavily criticised as they may promote arbitrary allocations of funds, and perpetuate existing inefficiencies and inequities in health services. In particular, many publicly financed health systems have used historically devolved budgets based on incremental adjustments to past expenditure and activity levels. This approach creates strong perverse incentives for health schemes to inflate expenditure levels to attract larger budget allocations in the future, whilst historically inefficient behaviour goes un-penalised, thereby promoting further inefficient behaviour. Inequalities in access are also perpetuated as health schemes with historically greater access - and therefore greater utilisation - are rewarded with greater funds, whilst health schemes with historically poorer access receive relatively lower funds.

The response to these criticisms has been the development of more scientific approaches to predict expected health service costs. These approaches have sought to provide accurate and robust estimates of expected health service costs that minimise the potential for perverse incentives when used to determine prospective budgets for health schemes. Scientific developments have focussed primarily on predicting costs to set capitated rates of reimbursement for health schemes. Capitation rates generally indicate how much a health scheme would spend on an individual if it were to provide some standard level of health services accounting for the influences of the individual’s health, demographic and social characteristics on health service expenditure.

The methods used to determine the influence of individual’s characteristics on health service costs, and to incorporate those influences into capitation payments are collectively known as risk
adjustment methods. A more technical definition of risk adjustment is the use of information to calculate the expected health expenditures of individuals over a fixed period of time to determine payment rates for health schemes in the pursuit of health system objectives (Van de Ven and Ellis 2000). The information used in risk adjustment methods seeks to capture the major determinants of health service expenditure.

In general, the determinants of health service costs/expenditure can be divided into two groups of factors: needs factors and supply factors. Needs factors may be defined as the demographic, health, and social characteristics of individual’s that give rise to interpersonal differences in the need for health services and therefore expected health services costs. Supply factors may be defined as the organisational, managerial and policy characteristics of health service providers that give rise to variations in expected costs unrelated to the needs of individual’s who receive services \(^2\). These provider characteristics reflect a range of potential sources of variations in costs. The main sources of variation in costs include variations in local clinical policy, managerial efficiency of providers, and variations in local availability of health services.

Supply side factors may also be divided into two groups: unavoidable and avoidable cost differences. Some aspects of efficiency and policy will be under the control of service providers whilst unrelated to the health service needs of individuals covered by health schemes. Supply side variations in costs that may be directly influenced by health schemes and that are unrelated to need may be termed avoidable cost differences. However, some aspects of efficiency and policy will lie outside of the direct control of service providers. These supply side variations in costs may be termed unavoidable cost differences. Unavoidable cost differences are generally due to variations in local factor input market conditions and in location specific costs of service provision. Local factor input market conditions may affect provider costs where local wages and rents vary according to the extent of local labour market conditions, availability of land etc. Location specific costs of service provision may affect provider costs where the physical location of health services and populations influence the costs of service delivery.

Given the vast array of alternative health system models in the world, the approach to risk adjustment adopted by different countries has been remarkably similar. The primary goal of all risk adjustment methods has been the prediction of the expected costs of health services for individuals within health schemes based on analysis of needs and supply factors. As such, all approaches to risk adjustment have shared significant similarities.

However, it is worth noting that the aims of risk adjusted capitation have differed slightly depending on whether capitation is to be used in a competitive health scheme model, such as the US and the Netherlands, or in a non-competitive health scheme model, such as the UK or Australia. (These implications of these differences for risk adjustment methods are discussed in section 6).

Under the competitive health scheme model health schemes are allowed to compete for enrollees under a range of possible financing and delivery arrangements, for example

\(^2\) This distinction has also been drawn by Carr-Hill et al (1994) and Rice and Smith (1999) – who use the terms legitimate and illegitimate factors – and Van de Ven and Ellis (2000) - who use the terms solidarity and non-solidarity factors. Whilst the terminology adopted elsewhere reflects the normative judgements required in determining factors relevant for capitation, we will distinguish between needs and supply factors to avoid what may be considered overly value laden terminology.
competitive health insurance, managed competition etc. A principle role of risk adjustment in this type of model is the promotion of efficiency in the operation of the insurance market (Rice and Smith 1999).

Risk adjustment seeks in this context to ensure that capitation payments are adequate to cover predictable future expenditures for enrollees to minimise the potential for adverse selection and cream skimming (Hutchison et al 1999).

Under the non-competitive health scheme model health schemes are not permitted to compete for enrollees, for example enrolled populations are determined by geographically defined administrative regions, employee insurance schemes etc. Incentives for adverse selection and cream skimming are therefore largely removed. A major role of risk adjustment under this model is the pursuit of explicit equity objectives (Rice and Smith 1999). Typically, explicit equity objectives have related to equal funding for equal need, or equal government subsidy for equal need across health schemes which individuals are compelled to join (Rice and Smith 1999).

Under all health scheme models, risk adjusted capitation funding models consist of four main elements reflecting the major determinants of health service costs (Peacock 1997):

(i) The number of persons covered under each health scheme,

with weights for health scheme enrollees to reflect:

(ii) their demographic related health care needs

(iii) their health status and social related health care needs

(iv) differences in the costs of service provision due to supply side factors.

The weights for supply side factors may be divided into weights to reflect:

(a) avoidable cost differences eg service provider efficiency, managerial and clinical policies

(b) unavoidable cost differences eg local factor input market conditions and location specific costs

The weights used in capitation funding models are derived from predictions of the expected costs of health services estimated using risk adjustment methods. All capitation funding models therefore employ a standard approach of calculating per capita rates based on individuals’ expected costs of health care. The prospective budget of an individual health scheme is found by aggregating the expected costs of health care for all individuals enrolled with that scheme. Differences between capitation models emerge only in the robustness of risk adjustment methods used to estimate expected health service costs, and in the choice of needs and supply factors that are used to predict costs.

The two key methodological issues in risk adjustment methods for predicting health service costs lie in the selection and estimation of weights for needs and supply factors in a capitation system. Selection of needs and supply factors relates to identifying the key determinants (predictors) of variations in health service costs, and estimation of weights for those factors relates to the magnitude of their impact on costs.
(i) Choice of Needs and Supply Side Factors

Rice and Smith (1999) characterise the capitation payment for a given individual as that person’s relative expenditure needs, and the characteristics to be taken into account in calculating those needs, as needs factors.

They suggest that needs factors should be chosen, ceteris paribus, on the basis that they represent material influences on the need to consume the health services in question. There have been a range of responses to the question of how to make such choices.

Rice and Smith (1999) distinguish between normative and empirical approaches to identifying needs factors. Normative approaches employ epidemiological and other scientific evidence to form judgements about the selection of needs factors. Essentially this has often involved synthesising available evidence on the link between key health and socioeconomic characteristics and need, to attempt to inform the choice of needs factors and weights for those factors by a panel of experts. As such, normative approaches have not formally employed empirical analysis to select relevant needs factors or determine appropriate weights for those factors, they have instead relied heavily on subjective assessment.

Empirical approaches are based on identifying statistical associations between potential needs factors and health service expenditure using models based on empirical data and evidence. These approaches have typically focussed on the use of historic data on health/socioeconomic characteristics and health service utilisation/expenditure to select needs factors and estimate weights for those factors using regression analysis. Current methods are based heavily in the empirical approach, although there may be varying degrees of a normative judgement implicit in empirical analysis.

Within the sets of weights used in risk adjustment a range of measures/variables included as needs and supply side factors in different capitation models have been identified (Hutchison et al 1999, Rice and Smith 1999, Van de Ven and Ellis 2000). Examples of these variables are shown in tables 2.1 and 2.2.

Table 2.1 indicates a wide range of needs factors employed across a range of risk adjusted capitation models in use. Needs factors have been grouped under 6 headings, although the distinction between some groups of factors may be somewhat blurred. The final group of needs factors refers to the extent to which an individual’s geographic location results in variations in the need for health services.

The treatment of supply side factors has been more troublesome than for needs factors, with a range of approaches adopted across different countries using some or none of the measures shown above. The two groups of factors shown in table 2.2 refer to potential sources of unavoidable cost differences (factors exogenous to health scheme decision making) and avoidable cost differences (factors endogenous to health scheme decision making). Some measures are frequently not directly observable, such as practice style, but may still be statistically modelled using recent advances in regression techniques.
<table>
<thead>
<tr>
<th>Category</th>
<th>Potential Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demography</strong></td>
<td>Age</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td>Disadvantaged ethnic groups - eg ATSI/Other, Maori/Pacific Islander/Other</td>
</tr>
<tr>
<td><strong>Employment/Disability Status</strong></td>
<td>Unemployment/Welfare Status – employed, unemployed, pensioner</td>
</tr>
<tr>
<td></td>
<td>Unemployment/Disability – temporarily unable to work, permanently sick</td>
</tr>
<tr>
<td><strong>Health Status</strong></td>
<td>Self reported/survey-based morbidity</td>
</tr>
<tr>
<td></td>
<td>Permanent disability/dependency status</td>
</tr>
<tr>
<td></td>
<td>Mortality rates (crude and standardised)</td>
</tr>
<tr>
<td></td>
<td>Previous inpatient/outpatient diagnosis</td>
</tr>
<tr>
<td></td>
<td>Low birth weight</td>
</tr>
<tr>
<td><strong>Socioeconomic Status</strong></td>
<td>Homelessness</td>
</tr>
<tr>
<td></td>
<td>Martial status/cohabitation</td>
</tr>
<tr>
<td></td>
<td>Income</td>
</tr>
<tr>
<td></td>
<td>Socioeconomic Status/Social class</td>
</tr>
<tr>
<td></td>
<td>Education level</td>
</tr>
<tr>
<td></td>
<td>Religion</td>
</tr>
<tr>
<td></td>
<td>Elderly Living alone</td>
</tr>
<tr>
<td></td>
<td>Housing tenure/quality</td>
</tr>
<tr>
<td></td>
<td>No carer in household</td>
</tr>
<tr>
<td><strong>Geographic location</strong></td>
<td>Region/area of residence</td>
</tr>
<tr>
<td></td>
<td>Rurality</td>
</tr>
<tr>
<td></td>
<td>Urbanisation</td>
</tr>
<tr>
<td></td>
<td>Remoteness</td>
</tr>
<tr>
<td></td>
<td>Population Density</td>
</tr>
</tbody>
</table>

Table 2.1  Examples of needs factors
### Table 2.2 Examples of supply side factors

<table>
<thead>
<tr>
<th>Category</th>
<th>Potential Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Provider Characteristics – exogenous to health scheme</strong></td>
<td>Input prices&lt;br&gt;Labour prices&lt;br&gt;Capital rental prices&lt;br&gt;Location specific service delivery costs – excess rural service delivery costs&lt;br&gt;Rurality&lt;br&gt;Urbanisation&lt;br&gt;Remoteness</td>
</tr>
<tr>
<td><strong>Provider Characteristics – endogenous to health scheme</strong></td>
<td>Market power - ability to get price discounts&lt;br&gt;Insurance coverage features - deductibles, co-payments etc&lt;br&gt;Contracting arrangements&lt;br&gt;Practice style&lt;br&gt;Health scheme management policies and practices</td>
</tr>
</tbody>
</table>

#### (ii) Needs Factor Weights

Following identification of the relevant needs factors (by whatever means), weights reflecting their relative impact on the need for health service expenditure must be determined. In practice this has been achieved almost entirely through empirical analysis. Rice and Smith (1999) found a significant range in the levels of refinement and statistical rigour in the 19 countries they studied, largely reflecting variations in data availability and quality. Findings from the 19 country survey are summarised in table 2.3.

Approaches have ranged from identifying the average expected health service expenditure for individuals with identified characteristics, for example age, sex, ethnicity etc. to regression models to explain observed variations in expenditure patterns. Typically the former requires individual level data, whilst the latter has been a response to a lack of individual level data through the use of aggregate level analysis. From the findings of their 19 country survey Rice and Smith (1999) describe two main approaches to determining weights for needs factors: the matrix approach and the index approach.

The matrix approach uses one or more needs factors to create a matrix, or contingency table, of expected annual health service costs for individuals with identified needs factors. Needs factors may include age, sex, ethnicity, morbidity, socioeconomic status etc. If, for instance, the matrix uses eight age, two sex, and two ethnicity categories it will have 8x2x2 cells (32 cells), each cell with an associated expected health service cost. This approach produces a large amount of detailed information for setting capitation rates, placing significant demands on the data. It requires not only that needs factors data be available at the individual level, but also that recording of such data is universal and reliable across all health schemes to avoid systematic
bias being introduced. Statistical or judgemental methods can be used to reduce the number of cells for which entries are required within the matrix. The matrix approach is used by several countries to cover a range of needs factors – France, Israel (age alone) Germany, Switzerland (age and sex) Sweden (age, sex and a range of socioeconomic characteristics). Other examples include the Netherlands, NZ and Medicare in the United States.

The index approach generally uses aggregate data on the characteristics of health scheme enrollees to construct aggregate indices of predicted expenditure. These indices are often referred to as needs indices, but do not reflect need in the conventional epidemiological sense. They are, instead, expenditure indices based on health service costs predicted by the characteristics of enrollees in the health scheme. The characteristics of enrollees are captured in terms of needs factors in risk adjustment methods, hence needs (based expenditure) indices.
Table 2.3  Capitation experiences in 19 countries (from Rice and Smith 1999)

<table>
<thead>
<tr>
<th>Country</th>
<th>Scheme</th>
<th>Health Schemes</th>
<th>Individual level factors</th>
<th>Scheme level factors</th>
<th>Other factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>New South Wales Resource Distribution Formula</td>
<td>17 Area Health Services (geography)</td>
<td>Age Sex Ethnic group Homelessness</td>
<td>Mortality Education level Rurality</td>
<td>Private utilisation Cross-boundary flows Cost variations</td>
</tr>
<tr>
<td>Belgium</td>
<td>National Institute for Sickness and Disability Insurance risk adjustment scheme</td>
<td>100 sickness funds (competitive)</td>
<td>Age Sex</td>
<td>Unemployment Disability Mortality Urbanisation</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>Alberta Population Based Funding Model</td>
<td>17 Regional Health Authorities (geography)</td>
<td>Age Sex Ethnicity Welfare status</td>
<td>Remoteness</td>
<td>Cross-boundary flows Funding loss protection Cost variations</td>
</tr>
<tr>
<td>Finland</td>
<td>State Subsidy System</td>
<td>452 municipalities (geography)</td>
<td>Age Disability</td>
<td>Archipelago Remoteness</td>
<td>Tax base</td>
</tr>
<tr>
<td>England</td>
<td>Resource Allocation Formulae</td>
<td>100 health authorities (geography)</td>
<td>Age</td>
<td>Mortality Morbidity Unemployment Elderly living alone Ethnicity Socioeconomic status</td>
<td>Cost variations</td>
</tr>
<tr>
<td>France</td>
<td>Regional resource allocation</td>
<td>25 regions (geography)</td>
<td>Age</td>
<td></td>
<td>Phased implementation</td>
</tr>
<tr>
<td>Germany</td>
<td>Federal Insurance Office risk adjustment scheme</td>
<td>Sickness funds (employment/competitive)</td>
<td>Age Sex</td>
<td></td>
<td>Income base</td>
</tr>
<tr>
<td>Israel</td>
<td>National risk adjustment scheme</td>
<td>4 sickness funds (competitive)</td>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>Regional resource allocation system</td>
<td>21 regional governments (geography)</td>
<td>Age Sex</td>
<td>Mortality</td>
<td>Damping mechanism</td>
</tr>
<tr>
<td>Netherlands</td>
<td>Central Sickness Fund Board risk adjustment scheme</td>
<td>26 sickness funds (competitive)</td>
<td>Age Sex Welfare/disability status</td>
<td>Urbanisation</td>
<td>Retrospective adjustments Income base</td>
</tr>
<tr>
<td>New Zealand</td>
<td>Health Funding Authority Population Based Funding Formulae</td>
<td>4 regions (geography)</td>
<td>Age Sex Welfare status Ethnicity</td>
<td>Rurality</td>
<td>Phased implementation</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>Health Board Allocation Formula</td>
<td>4 health boards (geography)</td>
<td>Age Sex</td>
<td>Mortality Elderly living alone Welfare status Low birth weight</td>
<td>Rural costs adjustment</td>
</tr>
<tr>
<td>Country</td>
<td>Scheme</td>
<td>Health Schemes</td>
<td>Individual level factors</td>
<td>Scheme level factors</td>
<td>Other factors</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------</td>
<td>----------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Norway</td>
<td>Local Government Finance System</td>
<td>19 county governments (geography)</td>
<td>Age</td>
<td>Mortality</td>
<td>Tax base</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sex</td>
<td>Elderly living alone</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td>Health Authority Revenue Allocation scheme</td>
<td>15 health boards (geography)</td>
<td>Age</td>
<td>Mortality</td>
<td>Rural costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>Regional resource allocation system</td>
<td>7 regions (geography)</td>
<td></td>
<td></td>
<td>Cross-boundary flows</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Declining population</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>adjustment</td>
</tr>
<tr>
<td>Sweden</td>
<td>Stockholm County hospital resource allocation formula</td>
<td>9 health care authorities (geography)</td>
<td>Age</td>
<td>Urbanisation</td>
<td>Phased implementation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Living alone</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Employment status</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Housing tenure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Previous inpatient</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>Federal Association of Sickness Funds risk adjustment scheme</td>
<td>Sickness funds (competitive)</td>
<td>Age</td>
<td>Income base</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>Medicare + Choice (from 2000)</td>
<td>HMOs (competitive)</td>
<td>Age</td>
<td>Labour costs</td>
<td>Phased implementation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Disability</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Welfare status</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Previous inpatient</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>County of residence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>Veterans Equitable Resource Allocation</td>
<td>22 networks (geography)</td>
<td>Dependency (x2)</td>
<td>Labour costs</td>
<td>Phased implementation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wales</td>
<td>Health Authority allocation formula</td>
<td>Health authorities (geography)</td>
<td>Age</td>
<td>Mortality</td>
<td>Sparsity cost adjustment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sex</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Needs based expenditure indices may consist of multiple measures for needs factors (e.g., Belgium and UK) which may then be combined into a single aggregate needs-based expenditure index (e.g., UK). For example, if health scheme populations are defined geographically and needs factors include age, sex, and ethnicity, then the weights on each of these needs factors may be applied to each area's population to determine its predicted expenditure. The index approach typically relies on aggregate level data, which allows use of a much wider range of potential needs factor variables, particularly where health scheme populations are defined geographically and coincide with census data boundaries.

In practice, the distinction between matrix and index approaches has been somewhat blurred. Several formulae use a mixed approach (England, NSW, Finland, and Italy) where age and sex...
are often used to construct a first matrix relating demographic characteristics to need. Indices are then developed for health and socioeconomic needs factors, standardised for age and sex. Weights from both approaches are then multiplied together to derive final predictions of per capita expenditure.

Adjustment for demographic needs has been the least controversial element of risk adjustment methods for predicting costs. It has been well established that age and sex lead to different needs for health services and place different demands on health service providers. Furthermore, many countries have collected detailed data on health service expenditure and utilisation, and their relationship to age and sex. As a result many countries have individual level data which is amenable for use in risk adjustment methods, and the majority of capitation schemes have therefore used a matrix approach to establishing demographic needs factor weights. Typically this has included the development of weights for between 7 and 18 age bands. The treatment of sex has varied between capitation schemes, however. Some countries have included sex as a needs factor as well as age and developed weights in a matrix format, whilst others have only developed weights for age.

Conversely, adjustment for health and social needs has been the most controversial element of risk adjustment methods, and technically the most demanding. This has resulted in significant attention in the literature on appropriate methods and their robustness. Rice and Smith (1999) identified six reasons why the selection of health and social needs factors has been considered to be controversial:

(i) relevant data are often in short supply;
(ii) research evidence on appropriate needs factors is sparse, dated or ambiguous in its implications;
(iii) there is great difficulty in establishing the extent to which a particular needs factor is independent of other needs factors, that is, in handling covariances between needs factors;
(iv) it is very difficult to disentangle legitimate needs factors from other policy and supply influences on utilisation;
(v) it is often difficult to identify the health care costs associated with a proven needs factor; and
(vi) the recipients of public sector budgets often feel they have a clear idea about which needs factors will favour their area, and so will seek to influence the choice of needs factors through the political process.

### 2.2 Criteria for Assessing Risk Adjustment Methods

A number of studies have identified a range of interrelated criteria for assessing alternative approaches to risk adjustment (Van de Ven and Ellis 2000, Rice and Smith 1999, Hutchison et al 1999). The criteria proposed across these studies are broadly consistent, and have been grouped under four headings: appropriateness of incentives, validity, feasibility, and acceptability.

(i) **Appropriateness of Incentives**

The appropriateness of incentives generated under risk adjusted capitation payments is the most important criteria over which risk adjustment methods should be assessed (Van de Ven and Ellis
There are many possible distortions or undesirable responses created by a given set of risk adjustment factors, and the extent to which these compromise the objectives of the capitation system is a pivotal issue evaluating alternative risk adjustment methods. These potential undesirable responses broadly relate to:

- the explicit selection of some patients over others (adverse selection);
- the oversupply of services to profitable patients, and the under-supply of services to unprofitable patients (cream skimming);
- the oversupply of services to patients and population groups with relatively low levels of need, and the under-supply of services to patients and population groups with relatively high levels of need (inequities in access); and,
- the distortion of information for payment purposes to funders (gaming the system).

The first three responses reflect the most common objectives of capitation systems – equity and efficiency - and will be applicable with varying degrees of importance depending on the particular context under consideration (e.g. whether the health system model is competitive or non-competitive). The fourth response is dealt with below under the criteria of feasibility.

Van de Ven and Ellis (2000) provide a succinct summary (see table 2.4) of the way in which health schemes may respond to the incentives created by risk adjusted capitation models. A further potentially undesirable response, both under competitive and non-competitive models, may relate to incentives for health schemes to cut back on those aspects of quality of care that are unobservable to reduce costs (quality scrimping). These potential responses are to be considered more fully in a later paper.

(ii) Validity

The most common criteria against which risk adjustment methods are assessed in practice is the validity of methods to predict future health expenditure, where validity is defined as the degree to which a measure actually assesses what it purports to measure. The reason for the focus on validity lies mainly in the relative ease with which the performance of statistical models may be evaluated. However, whilst appropriateness of incentives remains the most important criteria for assessing risk adjustment methods, validity may still provide useful information on the relative performance of alternative methods. Furthermore, if the objective of risk adjustment is simply to gain the most accurate prediction of future expenditure as is possible, then validity may become the most important evaluation criteria.

The most commonly used measure of the validity of risk adjustment methods is the predictive power of regression models used to select and weight needs and supply factors. This reflects the predominance of empirical (regression) techniques in risk adjustment. Typically the $R^2$ value for regression models has been used to assess predictive power. The $R^2$ value in individual level regression models measures the proportion of variance in individual expenditures that is explained by a set of individual and health scheme level, needs and supply factors. In aggregate level (small area) regression models it measures the proportion of variance in (small area) population expenditures that is explained by a set of (small area) population level needs and supply factors.
However, the predictive power of alternative risk adjustment methods must be interpreted with some caution if the objectives of capitation include efficiency and/or equity goals. Higher predictive power may not necessarily mean a “better” risk adjustment model if it creates inappropriate incentives (Van de Ven and Ellis 2000, Hutchison et al 1999, Rice and Smith 1999). For instance, a regression model that perfectly predicts future expenditure, based on current health service availability and utilisation patterns, will promote the replication of current and clinical practices and policies. Some of these practices and policies may be inappropriate in terms of efficiency and equity goals. Therefore, a regression model with lower predictive power, but which provides incentives consistent with the objectives of risk adjustment, may be preferable.

<table>
<thead>
<tr>
<th>Health scheme benefit features</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Deductibles or co-payments for selected conditions.</td>
<td></td>
</tr>
<tr>
<td>Lifetime or annual coverage limits.</td>
<td></td>
</tr>
<tr>
<td>Coverage of pharmaceuticals or other specific services.</td>
<td></td>
</tr>
<tr>
<td>Exclusions for pre-existing conditions.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Payment rate classes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Efforts to attract more profitable payment rate classes, including family or individual contracts, employee or retiree, specific geographic area.</td>
<td></td>
</tr>
<tr>
<td>Selection of relative premium rates by payment rate classes.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attraction/avoidance of profitable/unprofitable enrollees</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Denying coverage</td>
<td></td>
</tr>
<tr>
<td>Cancelling coverage</td>
<td></td>
</tr>
<tr>
<td>Selective advertising</td>
<td></td>
</tr>
<tr>
<td>Pre-enrollment screening</td>
<td></td>
</tr>
<tr>
<td>Selective enrollment and dis-enrollment counselling.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changes in services offered</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection of specialists to include/exclude from health scheme</td>
<td></td>
</tr>
<tr>
<td>Over-provision of services that attract profitable enrollees</td>
<td></td>
</tr>
<tr>
<td>Under-provision of services that attract unprofitable enrollees</td>
<td></td>
</tr>
<tr>
<td>Change of place of service to increase payments</td>
<td></td>
</tr>
<tr>
<td>Unnecessary provision of services to code a diagnosis</td>
<td></td>
</tr>
<tr>
<td>Change in timing of services to increase payment.</td>
<td></td>
</tr>
</tbody>
</table>

(iii) Feasibility

Feasibility may be split into three main considerations in developing risk adjustment models: the availability of data, resistance of data to manipulation, and administrative feasibility.
(a) Availability of Data

Rice and Smith (1999) outline a number of principles that should be applied to needs factors data on which risk adjustment is to be based. Needs factors should be based on individuals’ characteristics that are: universally recorded across schemes in receipt of funds, consistent, verifiable, free from perverse incentives, not vulnerable to manipulation, consistent with confidentiality requirements, and plausible determinants of service needs.

Typically, individual data suitable for risk adjustment is more readily found in health systems which have developed and used insurance claims in reimbursement (whether private or social insurance). The regulatory requirements for providers and health schemes to submit claims for patients under this approach have provided some very rich data sources for risk adjustment. However, the lack of a national identifier to link records in some countries (eg Australia) may severely limit the ability of funding agencies to develop a coherent approach to risk adjustment that provides appropriate incentives for providers. Without proper record linkage analysis of expenditures and needs factors cannot be carried out across the whole spectrum of health services. Appropriate incentives for efficiency and equity may then be limited to a subset of services, with incentives between broad service areas at best weak, and at worst inappropriate.

For example, pain management for cancer patients may be provided through a GP or a hospital based specialist. If the use of services by cancer patients in the primary and secondary sectors cannot be linked (i.e. there is no common identifier for each patient across primary and secondary sector databases) analysis under risk adjustment methods is necessarily partial. Analysis may be able to identify relationships between needs factors and use in each sector in isolation, but will mask potentially cost-effective linkages between sectors. If pain may be managed more cost-effectively in a primary care setting, analysis without linked data will not be able to fully establish the potential for (cost-effective) substitution between primary and secondary services. This may lead to “artificially” inflated reimbursement rates for the secondary sector, based solely on analysis of (higher) hospital based costs. Specialists will then have reduced incentives to seek out more cost-effective options for pain management through the primary sector.

In practice the lack of individual level data meeting these criteria has led to the use of aggregate level analysis in some countries, particularly where insurance claims for patients are not used. The criteria listed above, may severely limit the range of available data for individual’s characteristics, for example the only data that conform in the UK are age and sex. Aggregate level analysis presents its own problems, however. These include; limited data from national censuses which may quickly become out of date, inconsistency in data recording across schemes for utilisation and morbidity data, and the vulnerability to manipulation of some aggregate level data (eg unemployment statistics).

(b) Resistance to Manipulation

There is a need for analysis to identify needs factors that cannot be manipulated by the health schemes, or that create perverse incentives (needs factors should be exogenous to health scheme decision making). Needs factors that may be prone to manipulation by health schemes include historic expenditure, diagnosis-based measures and health survey measures. Historic expenditures are not commonly used as needs factors in risk adjustment, in part because of the
ease with which health schemes can manipulate such data (see below). Diagnosis and survey based measures may be more difficult to manipulate, but Van de Ven and Ellis (2000) outline a number of potential ways in which such data may be distorted by health schemes (see table 2.5).

**Table 2.5** Potential distortions in health scheme data reporting

<table>
<thead>
<tr>
<th>Changes in diagnostic coding/other claims information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-coding of diagnoses to more serious conditions</td>
</tr>
<tr>
<td>Proliferation of diagnoses</td>
</tr>
<tr>
<td>Fraudulent diagnostic coding</td>
</tr>
<tr>
<td>Coding of “rule out” diagnoses</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attempts to influence survey based health measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrollee coaching</td>
</tr>
<tr>
<td>Non-random enrollee sampling</td>
</tr>
<tr>
<td>Biased corrections for non-response</td>
</tr>
</tbody>
</table>

Manipulation of diagnosis based measures has received some attention in the literature. Rice and Smith (1999) point out that the new risk adjustment system to be put in place in 2000 for the US Medicare system relies heavily on capitation based on a Diagnosis Cost Group scale based on the diagnosis code of recent hospital episodes. Recent past utilisation and coding of diagnosis may be rather vulnerable to manipulation by providers in order to attract higher capitation rates in the future, for example through diagnosis creep. Carter, Newhouse and Relles (1990) found some evidence of distortions in diagnostic coding under the Diagnosis Related Group case payment system for Medicare in the US, but suggested distortions in reporting may not persist in the long run. However, no evidence is available that directly relates to the manipulation of data under risk adjustment models (Van de Ven and Ellis 2000).

**(c) Administrative Feasibility**

Administrative feasibility requires that data are available for all individuals covered by health schemes (and all potential enrollees) with relatively little effort required (and hence low administrative cost) for data collection (Van de Ven and Ellis 2000). Again such (individual level) data may be more common in insurance claim based health systems. For data to be readily available for use in risk adjustment it should be routinely collected, standardised and comparable across health schemes, of acceptable quality, and easily validated. Large and representative samples should exist on which risk adjustment models can be developed and calibrated before implementation. Time lags between data collection and its availability for use in risk adjusted funding mechanisms should be minimised to ensure data is up to date, and data should be updateable so that analysis and weights can be re-calibrated over time.

**(iv) Acceptability**

A key element of risk adjustment, often overlooked, is that it should be acceptable to key stakeholders. If risk adjustment methods and results are not accepted by even one key stakeholder group, the chance of risk adjustment being successfully implemented become small. Specific stakeholder groups will include the general public, patients, providers, health schemes,
funders and politicians. Acceptability concerns may span a wide range of considerations, often depending on the country and context within which risk adjustment is being developed.

Examples of acceptability considerations for the public and consumers may include whether risk adjustment penalises sub-populations who have historically under-utilised services; whether privacy of information requirements are sufficient, particularly for diagnosis based information; and whether risk adjustment based on certain characteristics, eg race, is ethical (in some sense). Provider and health scheme concerns may include whether risk adjustment is based on clinically meaningful distinctions; the extent to which clinical freedom may be limited; and the level of financial risk to be borne by the scheme. Funders and politicians concerns may include whether the risk adjustment model is transparent and simple; the level of financial risk borne by both schemes and funding agencies; and the extent to which incentives are consistent with broader social and economic goals. In broader terms, acceptability also requires that risk adjustment and capitation is understandable to all stakeholders – ie it should be transparent, conceptually simple, and credible with respect to the relationship between level of need and level of funding (Hutchison et al 1999).

2.3 Normative Approaches to Risk Adjustment

As discussed above, normative approaches employ epidemiological and other scientific evidence in the largely subjective selection of needs factors, whilst empirical approaches are based on identifying statistical associations between potential needs factors and health service expenditure. However, from their survey of 19 countries Rice and Smith (1999) found that almost all of the countries studied had capitation systems based largely on analysis of empirical data and analysis of existing patterns of health utilisation.

An example of the normative approach was provided by one of the earliest capitation models, the RAWP formula in England. The formula was developed by the Resource Allocation Working Party (RAWP) in the early seventies and covered the Hospital and Community Health Sector in the English National Health Services (Department of Health and Social Security 1976). The objective of the capitation model was to correct existing inequalities in the pattern of resource distribution, and to construct an allocation mechanism which was responsive to populations’ health needs. The conclusions of the RAWP group were that resources should be allocated to Regional Health Authorities (geographically defined health schemes) on the basis of population weighted for three needs factors:

- differences in the age structure between areas;
- differences in the health needs for health care between areas;
- and unavoidable geographical differences in the costs of providing services.

The RAWP group recommended weighting for morbidity by using condition-specific Standardised Mortality Ratios (SMRs) as a proxy for the differential needs based expenditure of populations. The formula assumed that the need for health care had a one-to-one relationship with SMR in the absence of empirical evidence to suggest otherwise. The allocations based on the final formula were used as targets, which Regions would converge on over a number of years. The assumption of a one-to-one relationship between SMR and need was heavily criticised following
the introduction of the formula, and significantly contributed to its replacement with an empirically based model in 1990.

Some current examples of normative approaches include the Italian and Scottish systems which rely to some extent on SMR as a needs factor without empirical analysis of the link between SMR and utilisation. However, the Scottish capitation formula is currently under review. Norway uses a mix of normative and empirical approaches.

2.4 Historic Expenditure Approaches to Risk Adjustment

At the simplest level predictions of future expenditure, at the individual or health scheme level, may be based on extrapolation of past expenditure levels. This form of historic budgeting has been prevalent across the world for many years, where health schemes receive the previous years funding allocation plus or minus some adjustment factor. Typically these adjustment factors relate to efficiency or cost containment targets on the minus side, and growth targets for rising demand or identified unmet need on the plus side. Whilst simple, transparent and easy to establish, this approach to funding has very serious limitations in terms of the most important criteria – appropriateness of incentives. Historic budgets will typically both perpetuate and promote existing inefficiencies and inequities in health services.

Under a historic budgeting framework high cost providers are given larger budgets than low cost providers. Whilst some of the reasons for high costs may be “legitimate”, eg due to greater case complexity, high costs may also reflect inefficient behaviour and inappropriate service provision. Therefore, high cost schemes will be rewarded with relatively large budgets in future periods, effectively rewarding inefficient behaviour. Similarly, the financial incentives for efficient providers to remain efficient become very weak. In the worst case, there are strong incentives for health schemes to artificially inflate expenditures in order to attract greater budgets in the future.

Inequalities in access are also perpetuated as health schemes with historically greater access - and therefore greater utilisation and expenditure - are rewarded with greater funds in the future. Health schemes with historically poorer access, and hence lower expenditure will receive lower funding in the future. Where there is excess demand for health services, individuals covered under relatively low access health schemes may face rising unmet need over time. Furthermore, the reason for the historically low access to services in those health schemes may be unrelated to the level of need in the population it serves.

For example, deprived urban areas may have historically low access due to the problems of attracting medical staff. Under historic budgeting these areas, which typically have the highest level of need for health services, will receive relatively less per capita than areas with high levels of access - eg more affluent urban areas – that have lower health service needs.

A second potential use of historic expenditure in risk adjustment methods is to use prior year expenditures of health services as a predictor of future expenditures in regression models. This approach has been examined most closely in the US and the Netherlands, where several studies have regressed health expenditure on age and sex, and prior year expenditure at the individual level, to examine the usefulness of prior year expenditure as a predictor of current expenditure.
These studies have shown simple correlation coefficients on prior and current year health expenditure data to be between 0.2 and 0.3 (Van de Ven and Ellis 2000). Regression analysis have similarly shown an incremental dollar spent in a given year “predicts” spending of 20-30 cents in the next year (Newhouse et al 1989, Van de Ven 1992, Ash et al 1998). In terms of statistical validity, regression $R^2$ values have been reported in the range of 0.06-0.10 in a number of studies where age, sex and prior year expenditures have been regressed on health expenditure (Van de Ven and Ellis 2000, Van Vliet and Van de Ven 1992, Ash et al 1998). This represents a substantial improvement over regression models that include only demographic variables (see below), indicating prior year expenditure is a strong predictor of current expenditure.

This suggests prior year expenditure is a potentially useful measure in determining capitation rates. However, whilst prior year expenditure data appears to hold reasonable statistical validity in regression models, it is a weak basis for risk adjustment in terms of its performance under other criteria.

Risk adjustment models have been developed in many contexts because data sources are not sufficiently developed to support a cost-based payment system (Van de Ven and Ellis 2000). Therefore, it may be unreasonable to expect that providers will be able to provide accurate historical health expenditure data at the patient level. This may be particularly true where health schemes have developed models of system delivery based on broad level contracts with providers that do not require detailed (patient level) disclosure of utilisation and expenditure data. However, in several countries, perhaps most notably the Netherlands, Sweden and Australia, large amounts of patient level data do exist.

More importantly, the use of prior year expenditure as a need factor may create perverse incentives. Variations in prior year expenditures between individuals may reflect inappropriate discretionary clinician practices. Capitation payments based on those expenditures would then continue to reward inappropriate practices. The use of average prior expenditures may also mask important relationships and create inappropriate incentives. For instance, providers may seek to select high prior expenditure patients who had an acute condition rather than those who had high prior expenditure due to chronic conditions. This would allow providers to generate greater profits in the subsequent year.

Finally, as with historic budgeting, by utilising historic cost data there are weak incentives for inefficient providers to improve efficiency levels. The strong link between historic expenditure and future revenue suggest that inefficiency will actually be perpetuated and promoted under such a model, although the extent to which inefficiency occurs may be reduced by limiting prospective capitation payments to only a proportion of total provider expenditure. Similarly, there are weak incentives for efficient providers to remain efficient, as inefficient providers will be seen to be rewarded with proportionally greater revenue.
3 Individual Level Risk Adjustment Methods

The shortcomings of normative and historic expenditure based approaches to risk adjustment have led to a significant emphasis on developing more scientific and robust risk adjustment methods in recent years. These developments have focussed on regression analytic techniques to predict health service costs using empirical data. Whilst slightly different regression techniques have been employed in different contexts, they have one common element: they provide estimates of the expected level of health service expenditure for individuals within a given population, based on a combination of needs and supply factors. The main distinction between alternative regression based risk adjustment methods has been in terms of whether analysis is based on individual (patient) level data or aggregate (small area) level data. Section 3 examines developments in individual level risk adjustment methods, and section 4 examines aggregate level risk adjustment methods.

The main developments in individual level risk adjustment methods have occurred in the United States and the Netherlands, where health schemes are allowed to compete for enrollees. Competitive health scheme models have primarily focussed on the use of risk adjustment methods to reduce the incentives for cream skimming in the health insurance market. The pervasive concern with cream skimming has led to a significant government sponsored research program in risk adjustment methods in several countries, most notably the United States and the Netherlands. This research has concentrated on the prediction of expenditure at the individual level using regression models. This section provides a brief overview of the most significant capitation systems and risk adjustment advances in the US and the Netherlands, and evidence on their performance. More detailed attention is given to the validity of risk adjustment models and needs factors used in individual level regression analysis.

3.1 Risk Adjustment Using Individual Level Analysis

(i) Risk Adjustment in the United States

The US health sector consists of a vast array of arrangements for the finance and delivery of health services. Several funding agencies, both private and public sector, have explored and implemented risk adjustment methods for capitation with varying degrees of success. The resulting diversity in approaches to risk adjustment therefore precludes a full discussion of all methods and capitation models in the US. Instead, attention here is focussed on the US Medicare risk adjustment model. This model provides the most important example of risk adjustment in a US competitive health scheme model, both in terms of the magnitude of the funds being distributed, and in terms of the sophistication of methods employed in determining capitation rates.

US Medicare is a statutory federal scheme covering the elderly (65 years and over) and End Stage Renal Disease patients, covering 13% of the population and accounting for 17% of expenditure. Beneficiaries are encouraged to join HMOs, which receive 95% of the adjusted average per capita cost (AAPCC), based until the end of 1999 on enrollee age, sex, county of residence, welfare status, whether they lived in nursing home, and local area health costs.
Newhouse has estimated that approximately 20-25% of health service costs are potentially predictable. That is, 75-80% of variation in costs is estimated to be due to random variation, and is not predictable. By implication, this component of cost variation is not open to manipulation by providers and health schemes, and requires large numbers of enrollees in health schemes to allow adequate risk pooling (see section 7.2). The AAPCC formula, however, has been shown to explain only 1-3% of variation in expenditure (Newhouse et al 1997, Giacomini et al 1995).

AAPCC data has also been the focus of some concern, particularly as it has tended to be excessively volatile. As a result, variation in the AAPCCs has been dampened by and blended with national average cost data. Furthermore, weights for risk adjustment factors in the Medicare matrix have been derived empirically from fee-for-service providers, rather than HMOs. A concern is that HMOs typically attract healthier patients, so weights based on fee-for-service expenditure may be excessively generous towards HMO patients.

The Medicare+Choice program, implemented in 2000, is to enhance the basis of risk adjustment and to address some of the shortcomings of the existing AAPCC formula. The development of the new risk adjustment system has been the response to significant criticisms that the previously used demographic variables were weak predictors of variations in health service expenditure (Anderson et al 1990, Ash et al 1989, Newhouse 1986). The Health Care Financing Administration has therefore sponsored a large risk adjustment research program. The primary aim of this research has been to develop indirect measures of patient health status from utilisation data, to attempt to avoid the perverse incentives generated under funding based on historic expenditure (see section 2.4). The three major results of this research have been the development of Diagnostic Cost Group (DCG), Hierarchical Coexisting Condition (HCC) and Ambulatory Care Group (ACG) models.

The new Medicare risk adjustment approach is based on a matrix design constructed using age, sex, welfare status and disability factors. Individual level data is used to construct relative weights from Medicare beneficiary claims. For males and females the factors used to construct the matrix are: 12 age bands, previous disability, Medicaid status, and Principal Inpatient Diagnostic Cost Group (PIP-DCG). For each cell in the matrix a weighting factor is derived, and the overall adjustment factor is found by adding up across cells in the matrix. The national average weight is given by 1.0, so that when weights are applied to the AAPCC for the county of residence, an individual’s capitation payment is found. The result of the new system is a more aggressive risk adjustment policy by Medicare, accounting for disability prior to Medicare qualification, qualification for Medicaid support, and disability/illness in the year prior to the current budgeting period (PIP-DCG adjustment).

The introduction of PIP-DCG adjustments has been of most interest to those seeking to develop risk adjustment methods. The PIP-DCG indicates the most severe category of inpatient diagnosis experienced by the individual over the previous year. An individual is allocated into one of the 16 PIP-DCG categories representing increasing severity of disability, with a risk adjustment factor applied according to which cell the individual is allocated to.

DCGs use broad categories of patient classification based on previous inpatient diagnosis over a defined period, prior to the current budget period, to predict expenditure. Diagnoses are grouped on the basis of expected future expenditure (Ash et al 1989), with patients allocated to their highest DCG in the period under consideration. The HCC model allocates patients to one of 100
disease based groups based on both inpatient and non-inpatient diagnoses over a specified period (Ellis et al 1996a). Under this approach all diagnoses are added together in determining risk adjustment factors, rather than allocating patients to the “highest” group as under DCGs. ACGs are similar to DCGs, but are based on non-inpatient diagnoses. Later work has focussed on integrating ACGs into the DCG framework (Fowles et al 1996, Weiner et al 1996). The results of this research has been to increase the explanatory power of risk adjustment factors and expenditure models from 1% in the demographic factors only models, to 5.4% when PIP-DCG is included and 8.6% when HCC is included (Ellis et al 1996a). Ambulatory Care Groups (ACG) used in casemix systems use fee-for-service data to inform adjustments for age, sex and diagnoses (Starfield et al 1991, Weiner et al 1991, Weiner et al 1996), and were found to predict 7% of variation in expenditure for adults (Fowles et al 1994).

Other government sponsored research has explored the use of survey based measures of health status in conjunction with prior utilisation data as predictors of expenditure (Newhouse et al 1989), the use of encounter-based risk adjustment (Ettner et al 1998), and a ‘carve-out’ system for high cost patients (Frank et al 1995, Maguire et al 1998). These approaches have shown some positive results, especially in the modelling of self-reported health status and variations in health expenditure.

Several concerns remain with the development of the new model. First, evidence suggests that capitation will never be able to remove selection bias completely from the US competitive health scheme model. Nor, for that matter will it be able to do so for any competitive model. Perfect capitation rates can never be estimated properly, measurement error will always be present. Therefore, there will always be opportunities for health schemes to avoid selecting patients where their knowledge/judgement indicates a particular individual’s expected health care costs will exceed the imperfectly determined capitated rate. Second, a feature of capitated models are the incentives for schemes to cut back on unobservable quality of care items in an attempt to reduce costs.

Third and most importantly, DCGs have two properties which may be undesirable in a funding model (also in common with DRGs). Schemes may have incentives to hospitalise patients for unnecessary lengths of stay, in order to raise their DCG factor for next periods capitated payment. Also, schemes have incentives to up-code patients to a higher DCG to obtain greater future payments (DCG creep). The extent to which health schemes will engage in this type of behaviour will depend on the extent to which longer than necessary lengths of stay and DCG creep are externally observable. At a more general level, the use of explicit historical use data provides broader ranging incentives for providers to inflate utilisation and costs, as under retrospective reimbursement. Past use data is endogenous, thereby providers may influence future capitation payments through current expenditure patterns. Essentially providers have strong incentives to inflate current costs, as those costs may lead to higher capitation rates in the next period.

Finally, despite testing a wide range of potential needs variables, including measures of demography, socioeconomic status, health status, disability, physiology, prior utilisation and diagnoses, no risk adjustment model for all US Medicare enrollees has predicted over 10% of the variation in expenditure. Whilst the introduction of other needs factors at the individual level can add to explanatory power (eg socioeconomic characteristics), Newhouse et al found the major gains in explanatory power were to be found through the introduction of variables for individual’s
past utilisation and health status (Newhouse et al 1997). However, one half of predictable variation in the models still remains unexplained (Hutchison et al 1999).

(ii) **Risk Adjustment in the Netherlands**

The Dutch Health system provides cover for two types of health service utilisation under largely compulsory insurance schemes - primary and short term care expenses, and exceptional medical expenses such as for long-term disability and serious illness. Only the former scheme includes risk adjustment in health scheme payments, primarily due to difficulties in predicting “catastrophic” health events and related health expenditures.

Primary and short term care is funded through earnings related contributions by individuals, with matching employer contributions. Revenues are allocated to 26 sickness funds (health schemes) via risk adjusted capitation payments. Approximately 10% of expenditures then have to be raised by the sickness funds through premiums paid by enrollees. The purpose of risk adjustment is to equalise the level of these premiums across all enrollees to avoid cream skimming incentives.

The risk adjustment mechanism is based on a matrix approach. Age and sex weights, using a total of 38 five year age/sex bands are first derived from historical expenditure data. These weights are then adjusted for five welfare/disability status categories (permanently sick head of household, employed, temporarily/partially able to work, unemployed with dependents and pensioner). Regional factors are then applied for degree of urbanisation (very high, high, medium, low and rural). Finally, differential weights are placed on hospital care, specialist care and other care. The regional and service factors therefore pick up variations in individual’s needs (over and above age/sex and welfare/disability status), as well as geographic variations in provider costs and the supply of services.

The relatively large number of factors incorporated into the capitation model result in a matrix with a large number of cells. With 19 age, 2 sex, 5 welfare/disability status, 5 urbanisation and 3 service type factors the matrix would consist of \(19 \times 2 \times 5 \times 5 \times 3 = 2850\) entries. Clearly this would make the weights matrix cumbersome to operationalise. The Dutch model therefore reduced the number of matrix entries by calculating the expected costs for age and sex and assuming that the impact of the welfare/disability factors are constant across broad age groups. These age groups are for 0-14, 15-34, 35-44, 45-54 and 55-64 years (some employment status variables did not apply to 0-14 and 65+ age groups). This had the effect of dramatically reducing the number of entries in the matrix.

However, only 3-4% of the variation in expenditure in the Netherlands has been found to be explained by the age, sex, region, and welfare/disability status needs factors (Lamers 1998a).

Partly in response to the relatively weak predictive power of the existing model, the Dutch government has sponsored a large risk adjustment research program. This research program, commented on more fully below, has represented one of the major sources of methodological advances in recent years, particularly in the development of needs factors based on prior diagnoses.
3.2 Validity and Predictive Power

The most common measure of the predictive power of risk adjustment models is the $R^2$ measure. The $R^2$ value, in individual level regression models, measures the proportion of variance in individual’s health expenditures that is explained by a set of needs (and/or supply) factors.

An important note on assessing the predictive power of any risk adjustment model is whether it performs well in terms of predicting expenditures outside of the sample originally studied. Whilst a high $R^2$ may indicate a good statistical model (in terms of the ability of needs factors to explain variations in expenditure), those findings may only apply for that particular sample studied. A key test of the robustness of a risk adjustment model is whether that model is a good predictor of expenditure outside of the initial sample used in modelling. This may involve drawing a further sample on which to test the model, consisting of a broadly similar population or sub population from the same time period, or testing the model on later years of data for the same population. Ideally, $R^2$ values should therefore be presented for the original sample used to construct the model, and for out-of-sample predictions. However, reporting of, and analysis and $R^2$ values based on out-of-sample predictions is rather infrequent.

To correctly interpret $R^2$, particularly in comparing studies and models, Van de Ven and Ellis (2000) suggest 6 key determinants of the $R^2$ value (and the total variation) should be considered:

(i) The type of service under analysis: predictive power and $R^2$ vary systematically with the type of service being analysed. Using the same set of risk adjusters for the same sample, Newhouse et al (1989) found an $R^2$ of 0.05 for inpatient care and an $R^2$ of 0.25 for outpatient care. Similar results were also found by Wouters (1991) who reported large differences between types of outpatient services using out-of-sample predictions - ranging from an $R^2$ of 0.40 for drugs to an $R^2$ of 0.005 for surgery. Van Barneveld et al (1997) also found much greater predictive power for long term care, reporting an $R^2$ of 0.56. Van de Ven and Ellis (2000) suggest $R^2$ values for acute care are more typically in the region of 0.15.

(ii) The population or sub-population under analysis: analysis of specific sub-populations may result in much greater predictive power in risk adjustment models. In particular, individuals with chronic conditions have been found to have a much greater level of predictable health service costs. Kronick et al (1995) found $R^2$ values between 0.29 and 0.51 when analysing Medicaid enrollees with entitlement based on disabilities. The interpretation of these findings is that those individuals with chronic conditions generate expenditures which persist over time related to their condition, whereas acute episodes lead to short term expenditures which may not persist over time. Whilst the incidence of an acute or chronic condition may be largely influenced by random factors, chronic conditions result in chronic needs and related expenditure which are much less likely to be volatile over time.

(iii) The variation in explanatory factors: variation in the determinants of expenditure will influence variation in the whole model. The proportion of predictable variation explained by needs and supply factors used as explanatory variables will depend on the extent to which variation in the determinants of expenditure is predictable and measurable. For instance, greater variation in input prices which is stable over time (eg differences between urban and rural areas) will increase total variation in the model and the $R^2$ value. But, greater variation in input prices which
is unpredictable (e.g., changes in market power) will increase total variation and reduce the \( R^2 \) value.

(iv) The length of the time period being predicted: the use of longer time periods in analysis smooths out some of the random fluctuations in expenditure data. Ellis and Ash (1989) report an \( R^2 \) of 0.0089 using monthly data and an \( R^2 \) of 0.04 using annual data, when predicting monthly costs.

(v) The level of medical technology: Van de Ven and Ellis (2000) suggest, albeit without any empirical evidence, that the \( R^2 \) value may reasonably be expected to increase over time, specifically in relation to the level of medical technology. They hypothesise that improvements in the level of diagnostic technology (particularly in relation to genetic disease linkage) may improve the prediction of health expenditures. Similarly, technological advances may prolong the lives of those with chronic conditions, and with greater proportions of individuals with chronic conditions expenditures will be more predictable.

(vi) The year of the data analysed: similarly, Van de Ven and Ellis (2000) suggest that the substitution of outpatient for inpatient services over time will lead to greater predictive power in analytic models. Furthermore, the increase over time of prescribed drugs, the proportion who are elderly, and the proportion who are chronically ill are associated with greater predictive power.

In a discussion of the predictive power of risk adjustment models an important issue is to what extent variations in expenditure can be explained in practice. This has been the focus of a series of articles using individual level analysis of health expenditures based on US and Dutch data. Predictive power at the individual level can be examined through two approaches - models that estimate variation between individuals (cross-sectional data) or models which estimate variation within individuals (time-series data). The former has received attention from McCall and Wai (1983) and Newhouse et al (1989) who report upper bounds of 0.15 and 0.145 for the \( R^2 \) value respectively. The latter has been the subject of studies by Newhouse et al (1989, 1996) and Van Vliet (1992) who found predictable proportions to be 0.20 and 0.174 respectively (adjusting for the effects of autocorrelation). Van de Ven and Ellis (2000) therefore suggest the between and within individual evidence is consistent with the notion that risk adjustment regression models are capable of explaining approximately 20% of the variation in health expenditures. However, they suggest this may be a conservative estimate of the potential predictive power of individual level analysis because other predictive factors, not reflected in past expenditure (e.g., pregnancy, a recent diagnosis, etc.) may later become available, and used by health schemes to predict expenditure.

As discussed above, \( R^2 \) values may vary according to a number of factors. Newhouse (1989) reports maximum \( R^2 \) values for inpatient care of 0.08 and for outpatient care of 0.48 which are consistent with Dutch findings (Van Vliet 1992) which also found an extremely large \( R^2 \) of 0.80 for GP consultations and prescribed drug expenditure. Van de Ven and Ellis (2000) suggest this last finding implies that models excluding prescribed drugs may not be readily compared with models including prescribed drugs.

Van Vliet (1992) also found that predictive power increases for older age groups, supported by Newhouse (1989, 1993) who found the maximum \( R^2 \) for 14-64 year olds for outpatient services was 0.48, whereas the \( R^2 \) was 0.37 for 3-13 year olds. Van de Ven and Ellis (2000) suggest that
because of differences in the level (and also rate of change) of medical technology maximum $R^2$ values may not be comparable across countries, and that maximum predicted power should be estimated for individual countries for benchmarking the performance of risk adjustment models. The year of analysis may have an impact on the maximum $R^2$ value possible, though evidence is somewhat inconclusive. The studies reporting a maximum value of around 20% were based largely on data from the 1970s and 80s. A more recent study, using data from 1992-96 found a maximum $R^2$ value of 0.33 (Lamers 1999b), which may support the hypothesis of predictive power rising over time.

Finally, Van de Ven and Ellis (2000) identify three technical issues with the use of the $R^2$ value as an indicator of the validity of risk adjustment models: truncation of the dependent variable, logarithmic transformations, and overfitting models in small samples. These issues suggest a degree of caution is required in interpreting $R^2$ values, particularly when comparing risk adjustment models between different studies, different populations and different countries.

Truncation of the dependent variable in expenditure analyses is a relatively widely used data manipulation, where extreme - outlying - observations are omitted from analysis. Whether omission of high cost outliers is appropriate depends on the operation of the capitation payment system, for instance it is often justified on the basis that health schemes may reinsure very high cost patients, and therefore not bear their full cost (Van de Ven and Ellis 2000). Van de Ven and Ellis (2000) found that evidence from several studies was consistent in indicating a 30-70% increase in predictive power as a result of truncating the expenditure variable for a range of alternative regression models and needs factors. Log transformations of the data are also common for regression analysis but may inflate $R^2$ values because of the transformations to variables, and hence variation in the regression model. Similarly, small sample sizes have been shown to overfit regression models (systematically overstate $R^2$ values) (Ellis and Azzone 1998). Effects may even occur at sample sizes of 50,000, where $R^2$ values may be inflated by up to 30% compared to models estimated on one million observations.

### 3.3 Validity of Needs Factors

#### (i) Demography

The most basic, and most readily available, data for risk adjustment models are age and sex. Large differences exist between expenditure by individuals in relation to their age and sex. For example, Van de Ven and Ellis (2000) report over a ten fold difference in average expenditure on males aged 60-64 versus females aged 5-9 in studies of US and Dutch insurance schemes, and in US Medicaid enrollees. In dollar terms the figures range from approximately US$4,100 pa for males aged 60-64 to US$350 pa for females aged 5-9.

Age and sex have the advantages that they are easy to collect and document for use as needs factors in determining risk adjusted capitated payments. They are also generally considered to be acceptable to most stakeholders (Van de Ven and Ellis 2000). For example, age and sex adjustments on health insurance premiums are widely used across the world, whereas adjustments for other needs factors are not. Furthermore, age and sex data is independent of the health system, so is not readily gamed, or manipulated by providers. The major drawback with
age and sex as need factors is that they are relatively weak predictors of health expenditure at the individual level, predicting only 1% of individual level variation in expenditure (Van de Ven and Ellis 2000, Rice and Smith 1999).

(ii) **Diagnosis-based needs factors**

Over the last 20 years considerable research effort has been focussed on the development of diagnosis-based measures of need drawn from historical insurance claims data first in the US, and more recently in the Netherlands. This research has concentrated on developing predictors of future expenditure that do not create the perverse incentives for efficiency and equity that are created when prior utilisation data is used to determine prospective capitation payments. The development of diagnosis-based measures has led to the emergence of several classification systems which have been employed to varying degrees in risk adjustment and capitation systems. The three major classification systems are the Ambulatory Care Group system (ACG), the Diagnostic Cost Group system (DCG) (incorporating Hierarchical Condition Categories), and the Disability Payment System (DPS).

These systems have a number of common characteristics (Van de Ven and Ellis 2000). All diagnosis-based measures have been based on the notion that diagnoses will (at least in part) predict health care expenditures. This has led to groupings of diagnoses to create the alternative classification systems, based on the ability of groupings of diagnoses to predict future or current expenditure patterns. These groupings have been based on aggregation of ICD codes. With approximately 15,000 ICD codes it would be intractable to estimate and calibrate a risk adjustment model at the individual ICD code level as patient numbers in each ICD code would be very small, and the number of observations required very large as a consequence. In general, diagnosis data has been drawn from clinician diagnosis of conditions, rather than data on diagnosis from laboratory, diagnostic testing and medical supplies claims data is less reliable (Van de Ven and Ellis 2000). Regression analysis of variations in expenditure in relation to diagnosis groups have then been used to form the basis of weights for capitation payments at both the individual and health scheme level.

ACGs are based on aggregation of all ICD9-CM codes into 32 diagnostic groups (Weiner et al 1991, 1996) using ambulatory diagnoses only. These diagnostic groups can then be used in a number of alternative combinations, providing up to 83 mutually exclusive ACGs into which any given individual may be classified.

Most ACGs relate to broad diagnosis groups, such as acute: major or chronic medical unstable, but some relate to specific conditions, such as asthma. Thirteen ACGs are based on multiple counts under each of the initial 32 diagnostic groups, giving greater weight to patients with multiple conditions. Recent developments in the ACG classification system have led to the incorporation of inpatient as well as ambulatory diagnoses, but developments to date have been rather limited.

DCGs (Ash et al 1989, 1998, Ellis et al 1996a, 1996b, Pope et al 1998a, 1998b) have received the most attention of all classification systems. Early versions of this system were based on simple hierarchical models of diagnosis grouping, where modelling was used to identify a large number of clinically homogenous groups which were then aggregated into between 9 and 20 Diagnosis Cost Groups. Aggregation into the small number of DCGs was done on the basis of empirically determined similarities in the expected future costs of individuals with different primary inpatient
diagnoses. Individuals with multiple inpatient diagnoses were assigned to the most expensive DCG.

More recent developments in the US have significantly extended the DCG framework to include secondary inpatient diagnoses, outpatient diagnoses and other diagnoses assigned by clinicians. The simple hierarchical approach has also been modified so that hierarchies are only imposed on diagnostic groups when they are clinically related, by using information about multiple conditions. Development with clinicians has also led to a more detailed initial set of diagnostic groupings (543 groups), which are aggregated into 118 groups called Hierarchical Condition Categories (HCCs). These HCCs have been used in regression analysis of cost variations in the place of broader DCGs to estimate the incremental costs of specific conditions. Concerns about incentives for providers to manipulate diagnosis coding and admissions have been incorporated in modelling through the exclusion of selected HCCs.

Developments in the Netherlands have included expanding the DCG framework to include multiple years of hospitalisations (Lamers and Van Vliet 1996). The purpose of this approach was to capture the ability of diagnoses in several prior years to predict current expenditures, and to give greater payments to providers for certain diagnoses during several prior years so that payments for the chronically ill will more closely match actual expenditures. This work has suggested some reasonable gains in the predictive power of regression models from the inclusion of several prior years diagnoses, with a doubling of predictive power when 3 year prior DCGs were added to demographic variables in regression modelling (Lamers and Van Vliet 1996).

A major advantage of the DCG system is in terms of data availability, with large samples often available on which models can be estimated, and which provide good predictive accuracy (Van de Ven and Ellis 2000). However, a diagnosis-based approach to needs factors will frequently reward providers that more actively encourage individuals to seek treatment. Providers that do not directly treat, or admit, an individual, focussing instead on preventive or less “aggressive” treatment protocols may be relatively disadvantaged. These incentives become stronger if DCGs are used only for inpatient diagnoses, placing strong incentives on providers to seek out and admit greater number of patients.

Furthermore, the use of DCGs may encourage up-coding of patients on the margins between groups (“DCG-creep”) to receive greater capitated payments (in the same way case payment is prone to DRG-creep).

The DPS (Kronick et al 1996) was developed for US Medicaid disabled enrollees, and is based on similar principles to the DCG/HCC model. All diagnoses from clinical encounters are used within a hierarchical system for conditions. However, the DPS is more additive than the DCG/HCC system in its methods to account for the number of conditions an individual has within certain body systems. The generalisability of the DPS is unclear, however, as its development and application has focussed on people with disabilities (Van de Ven and Ellis 2000).
(iii) **Prescribed Drugs**

A second source of needs measures based on prior utilisation is the use of prescription drugs. Early work focussed on classifying drugs into different therapeutic classes and counting drug orders in each class (Hornbrook et al 1991). This approach was extended to include clinician judgement of severity to form weighted disease scores based on outpatient pharmacy utilisation and condition severity to develop the Chronic Disease Score (CDS) (Von Korff et al 1992). In conjunction with age, gender and health service visit data, the CDS was found to predict both hospitalisation and mortality. A later revision of the CDS re-calibrated the weights using empirical data for individual drug classes for 28 different conditions (Clark et al 1995), which resulted in an increase in variation in expenditure explained from 3% to 10% (with age and sex included in the model).

More recent work in the Netherlands has focussed on developing a pharmaceutical classification system based on homogenous cost groupings, resulting in 6 Pharmacy Cost Groups (PCGs) (Lamers 1998b). PCGs are based on the CDS system, using alike conditions and empirically determined similarities in future costs to construct the 6 broad groups. Whilst PCGs may have good explanatory power in regression models, concerns remain about the incentives created (Lamers 1998b). PCGs (and the CDS) have very large coefficients in regression models, resulting in the associated capitation adjustment far exceeding the costs of prescribed drugs used which are used in the regression model. Ellis (1985) reports that every dollar of drug expenditure predicts $3.73 of the next years health expenditure. Clearly then, there are very strong incentives for providers to inflate prior drug expenditure to attract greater capitation payments in subsequent years.

(iv) **Self-reported health information**

Self-reported survey information on health status has several advantages over prior utilisation and diagnosis-based information (Van de Ven and Ellis 2000). Health survey information can be obtained without contact with medical providers, no prior medical or insurance history is required; reflects individual’s perceptions of their own needs and expected utilisation, is uniform across health schemes and providers; and other relevant data, such as socioeconomic measures etc., may also be collected.

There are also disadvantages to self-reported measures of health; surveys are costly to undertake; response rates may be low and affect the robustness of empirical analysis, requiring large sample sizes; responses may be correlated with medical risk; reporting may be inaccurate due to errors or confidentiality issues. Furthermore, provider assistance in conducting surveys may be required, which may lead to problems in follow up and non-random sampling.

Van de Ven and Ellis (2000) present information on the validity of regression models that have sought to explain variations in health expenditure in terms of health status measures (see table 3.1). These studies show that adding self reported health status measures to demographic variables significantly increases the predictive power of risk adjustment models. However, self reported health status measures do not outperform prior diagnosis variables in explaining variations in health expenditure.

<p>| Table 3.1 | Comparison of R² Values |</p>
<table>
<thead>
<tr>
<th>Study</th>
<th>Newhouse et al 1989</th>
<th>Van Vliet and Van de Ven 1992</th>
<th>Fowles et al 1996&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Pope et al 1998&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Lamers 1999&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample Population</strong></td>
<td>US privately insured</td>
<td>Netherlands</td>
<td>US HMO enrollees</td>
<td>US Medicare</td>
<td>Netherlands sickness fund</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>N=7,690</td>
<td>N=20,000</td>
<td>N=5,780</td>
<td>N=10,893</td>
<td>N=10,570</td>
</tr>
<tr>
<td><strong>Age/Sex</strong></td>
<td>0.016</td>
<td>0.028</td>
<td>0.058</td>
<td>0.007</td>
<td>0.038</td>
</tr>
<tr>
<td>All socioeconomic&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.037</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Functional status&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Self reported chronic conditions&lt;sup&gt;a&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self reported health&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-Form 36&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>0.111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior year spending&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.064</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehensive survey&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.114</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis based&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.045</td>
<td>0.124&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.073&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.080&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>All Variables&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.090</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:  
<sup>a</sup> all models include age and sex as well as variables shown  
<sup>b</sup> Dependent variable was truncated at $25,000, which inflates $R^2$  
<sup>c</sup> ACG model  
<sup>d</sup> DCG/HCC model  
<sup>e</sup> Three-year DCG-model  
Source: adapted from Van de Ven and Ellis 2000, table 3
(v) **Mortality**

In individual level analysis mortality has been proposed as a potential explanatory variable in regressions on expenditure due to the excess costs incurred prior to death. However, whether mortality is a useful predictor of variations in expenditure at the individual level is not clear. One argument is that the costs associated with death are largely unpredictable, and have been found to be 250% above predicted (Van Vliet and Lamers 1998). The same study also found that the redistributive effect of including mortality as a needs factor in capitation would be very small (ie providers with large excess mortality would receive very small extra amounts of revenue even when mortality was explicitly included in modelling). A more effective means of reimbursing health schemes for the costs of death may be to control for deaths in the regression model (through dummy variables for patients who died during the period), and reimburse schemes retrospectively for those individuals.

Van de Ven and Ellis (2000) have questioned the political and social acceptability of increasing payment rates to health schemes with higher mortality rates. Mortality data may also be poorly coded and only partially available in some contexts, and may raise some concerns about privacy.

(vi) **Employment/Disability Status**

Several studies have found disability and functional health status to be relatively good predictors of future health expenditures at the individual level (Thomas and Lichenstein 1986, Hornbrook and Goodman 1996). In two studies the disabled and functionally impaired were found to have approximately double the expenditure of the unimpaired, and had strong predictive power after demography and prior utilisation were added to regression models (Lubitz et al 1985, Gruenberg 1989). Several countries also adjust for employment status in capitation payments, including the Netherlands.

Both employment and disability status have the advantage that (in most countries) they are universally recorded and frequently updated. However, neither set of measures are specifically designed to capture variations in health service needs, and are prone to systematic manipulation and misreporting. Perhaps crucially, employment and disability status are particularly poor measures of need in the elderly population, leading to the potential for systematic bias against this population group under capitation funding.

4 **Aggregate Level Risk Adjustment Methods**

The main developments in aggregate level risk adjustment methods have occurred in the UK, where health schemes are not allowed to compete for enrollees. Non-competitive health scheme models have focussed on the use of risk adjustment methods to reduce geographic inequities in access to services. The concern with equity and access has led to a significant government sponsored research program in risk adjustment methods in the UK and other, mainly European countries. This research has concentrated on the prediction of expenditure using small area regression analysis. The use of aggregate data has largely been a response to a lack of appropriate individual level data in some countries where individual level health insurance claims data is not required in the operation of the health system. This section provides a brief overview
of the most significant capitation systems and risk adjustment advances in the UK and evidence on their performance. More detailed attention is given to the validity of needs factors used in aggregate level regression analysis.

4.1 Risk Adjustment Using Aggregate Level Analysis

(i) Risk Adjustment in England and Wales

Much of the seminal work in developing aggregate level analysis for risk adjustment was undertaken in England in the late 1980s. In 1985 the Government announced a Review of the Resource Allocation Working Party (RAWP) formula. The aim was to ‘fine tune’ the formula to be more responsive to needs (Department of Health and Social Security 1986, 1988), and it was on the basis of this review that a new formula was introduced in 1990 and remained in force until 1994. The Review was conducted by constructing an empirically based model to analyse the determinants of hospital expenditure in small areas (electoral wards) (Coopers and Lybrand 1988). By explaining variations in expenditure hospital and community health service patterns across England the authors sought to find the underlying health needs which drive health service costs. The Review of RAWP recommended that the new formula should use all cause SMR as a measure of needs, the weight on SMR should be reduced, and that the formula should also include the Jarman index of social deprivation. The Government chose to only partially implement these recommendations, omitting the Jarman index and changing the weight on SMR. A fuller review of these methods has been given by Mays and Bevan (1987).

More recent work in the UK has focussed on statistical methods for the selection of ‘appropriate’ needs factors to address shortcomings identified in the Review of RAWP methods. The “York formula” (Carr-Hill et al 1994a, 1994b, Smith et al 1994) was developed by identifying needs factors that were statistically significant in explaining actual expenditure patterns. The York formula consists of four elements: population, an age adjustment, a health and social needs adjustment, and a unavoidable cost differences adjustment. The budget for a health scheme under the York formula can be expressed as:

$$\text{Budget} = \text{PerCap} \times \text{Pop} \times (1+a) \times (1+n) \times (1+c)$$

PerCap represents the national average per capita expenditure on hospital and community health services.

To calculate the budget for a health scheme PerCap is multiplied by Pop, the number of enrollees in the scheme. This figure is then weighted for the age structure of the population (the age adjustment a), the health and social needs of the population over and above age considerations (the needs adjustment n), and unavoidable differences in the local costs of providing health care (the relative cost factor c). The factors a, n, and c reflect the impact of age, health and social needs, and local supply side factors on expenditure, and are estimated by regression analysis. The national averages of a, n, and c are zero. Therefore, the formula predicts expenditure for individuals enrolled in each health scheme based on needs and supply side factors, and the budget for each scheme is calculated by aggregated predicted expenditure across all of its enrollees.
Under this approach each of the three adjustments are treated independently. An area can have a relatively young population, leading to a negative value of \( a \). At the same time the area may also experience high morbidity, leading to a positive value of \( n \). Finally, depending on local labour and capital costs, the area may have either a negative or a positive cost adjustment factor \( c \). Therefore, any one health scheme can have some parts of the formula working to increase its revenue share, whilst other parts serve to reduce it. For example, consider a health scheme with a relatively young population, for which per capita health service needs are estimated to be 4% below the national average. This leads to an age weighting of 0.96. However, given its age structure, the scheme has relatively high health and social needs, with morbidity 11% above the national average. This leads to a needs weighting of 1.11. Finally, the scheme is located in a part of the country where the cost of purchasing a given package of health care is estimated to be 15% above the national average. This leads to a cost weighting of 1.15. The approximate net effect of these three considerations is to weight the scheme’s total population by 0.96 \times 1.11 \times 1.15 = 1.225. That is, for every person in the population, the scheme will receive about 22.5% more than the national average per capita allocation.

The risk adjustment analysis undertaken for the York formula was commissioned to address several methodological and statistical shortcomings identified in earlier formulae (Mays 1989, Sheldon and Carr-Hill 1992). The empirical study was based on the notion of identifying a national “average” response (in terms of inpatient utilisation/expenditure) to variations in health care needs. Two types of determinant of demand were considered important in causing utilisation: the health care needs of the population, and the supply of health care facilities. A wide range of potential indicators of need factors were constructed, including indirect social determinants of demand for health care as well as direct measures of health status. The consideration of supply effects reflected the widely held belief that the availability of health care services affects the demand for those services. Potentially this occurs in two ways: firstly, in situations of excess demand utilisation may be suppressed due to constraints in supply; and secondly, there is evidence which suggests that the supply of physicians can induce demand (Cromwell and Mitchell 1986). The intention was therefore to construct a statistical model relating utilisation to indicators of health care needs and supply. A full description of the modelling techniques employed is given elsewhere (Carr-Hill et al 1994a, Carr-Hill et al 1994b). The theoretical model employed is shown in Figure 4.1.

**Figure 4.1  The York Model of the Demand for Health Care**

![Figure 4.1 The York Model of the Demand for Health Care](image-url)
The model suggests the level of utilisation in an area is determined by the availability of health care to that area, and the health and social needs of the population. The supply of health care to an area is determined by needs and utilisation (since earlier allocation mechanisms distributed funds on the basis of needs and utilisation), and other socioeconomic characteristics of that area which may shape the pattern of health care provision. The study recognised that health needs influence utilisation both directly (arrow a), and indirectly (arrow c), and allowed for the full effects of needs to be captured in modelling.

Mathematically the model can be represented by a system of two equations:

\[ U_i = f^1(N_i, S_i) \]  \hspace{1cm} (1)  
\[ S_i = f^2(N_i, U_i, X_i) \]  \hspace{1cm} (2)

Where the subscript \( i \) denotes the \( i \)th small area, \( U \) is utilisation, \( N \) are health care needs (including health characteristics and broader demographic and socioeconomic influences), \( S \) is the supply of health care\(^3\), and \( X \) are other socioeconomic characteristics (eg the residential desirability of an area). This mathematical model forms the basis for statistical analysis which relates inpatient utilisation to indicators of health care needs and supply using small geographical areas as the unit of analysis. Coefficients in equation 1 are then used as the basis of the needs adjustment \( n \) in the York formula. The methods used in the York formula provided three key innovations in terms of the treatment of simultaneity in the model, estimating the full impact of needs variables, and modelling policy influences.

In order to select needs factors to be included in capitation, and to estimate the magnitude of their associated weights, equation 1 was estimated. The aim of estimating equation 1 is to identify the key needs based determinants of variations in patterns of utilisation, whilst accounting for the influence of supply on utilisation. However, estimation of equation 1 through orthodox regression techniques, such as ordinary least squares, will produce biased model coefficients due to the presence of endogeneity in the model.

Endogeneity arises because the variables \( U \) and \( S \) are determined within the system of equations in the model, and \( S \) will be correlated with the error term in equation 1. To address this problem equation 1 was estimated using two stage least squares - an instrumental variable technique - which produces unbiased coefficients in the presence of endogenous variables.

The modelling approach adopted was to start with a general model, containing a large number of potential needs and supply variables, and to use statistical testing to incrementally move towards a smaller, more parsimonious, model which contains key needs determinants of variations in utilisation (following a ‘general to specific methodology’ (Hendry 1987) The rationale for this approach was that there are no incontrovertible reasons for the a priori exclusion of any socioeconomic variables, requiring that every needs variable should be considered, but only those which are highly correlated with utilisation be retained in the final model. The general to specific modelling approach also represented good statistical practice as it excluded variables

\(^3\) Supply was measured in terms of indices of the availability of hospital services, GP services, nursing home services, and private health services. Indices were based on gravitational models incorporating the physical availability of health services (eg beds, GPs), distance travelled by patients, and the effects of “competition” for health services from neighbouring populations.
which exhibited high degrees of collinearly or only weak associations with utilisation, and resulted in needs based expenditure indices which are not too large or complex. Modelling therefore started with a large model, and tested various functional forms (e.g. linear, multiplicative etc.) with the multiplicative model performing best under statistical tests of model specification. Supply variables were found to be endogenous under statistical testing, supporting the simultaneous equations approach.

The two stage least squares modelling process identifies legitimate needs factors which had an unambiguous statistical relationship with utilisation, but the coefficients on the supply variables reflect both health care needs and other socioeconomic characteristics which are not related to needs. The task was then to isolate the effects of needs on utilisation, both directly (arrow a in figure 4.1) and indirectly through supply (arrows b and c). This was achieved by running a final ordinary least squares of utilisation on the needs variables identified in the parsimonious model found by the process outlined above. The coefficients in this regression capture the full impact of needs on utilisation (Smith et al 1994). This approach allows for the fact that variations in existing supply to some extent reflect variations in legitimate health care needs.

The final stage of modelling was to re-estimate the parsimonious model using multilevel modelling techniques (Paterson and Goldstein 1991). A simple example of the use of multilevel models can be drawn from education, where the technique was developed. Analysis of student performance in exams may be studied through analysis of their test scores. A model may be estimated that assesses student performance at the individual (student) level by regressing their test score on individual level characteristics, such as age, sex, race and previous test results. This type of model would estimate the strength of relationships between key student characteristics and their performance. However, such estimates are likely to be biased (incorrect) because of other systematic effects on student performance. Student test scores will also be systematically related to their class and teacher, their school, and the state/education authority the school is in, through educational practice and policies. Multilevel modelling can directly estimate the impact of practices and policies at the class, school, and state/authority level, and produce unbiased estimates of the relationship between student characteristics and performance. The key issue is that performance at the individual level is systematically related to effects of the hierarchical system of education organisation and delivery.

In much the same way, patients are generally treated within wards or clinics which are part of hospitals or practices, which are part of health schemes/health authorities etc. The hierarchy of health service organisation and delivery means that individuals’ health expenditures will in part be determined by the characteristics of the individual, but also by the policies and practices found at different levels of the hierarchy. Just as in education, multilevel modelling techniques are required to provide correct estimates of the nature of these relationships (Rice and Jones 1997).

More formally, a fundamental assumption of the simultaneous equations modelling approach as described above is that the residuals from the model are independently distributed. However, it is quite possible that systematic variations in utilisation will occur between areas because of the policies and practices of different administrative regions (e.g. state health departments, regions within states). Multilevel modelling techniques take account of such clustering in the data, producing unbiased model coefficients (Paterson and Goldstein 1991). Multilevel models decompose variance in the model into two or more elements. The York formula methods used a two level model, which distinguished between variance at the small areas (electoral wards) level,
and at the Health Authorities (health scheme) level, producing estimates of the slope and intercept terms in equation 1 specific to each Health Authority. Results from the multilevel model were then used to construct a needs based expenditure index based on national average responses to needs in terms of utilisation, purged of the effects of different administrative regions.

The analysis produced a new formula based on two separate elements: the acute model, and the psychiatric model. Each contains SMR and a cluster of census-based needs variables which were found to be most important in the empirical analysis. Table 4.1 shows the health and social variables used in the York formulae.

<table>
<thead>
<tr>
<th>Table 4.1</th>
<th>York Acute and Psychiatric Models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acute Model Variables</strong></td>
<td></td>
</tr>
<tr>
<td>Standardised Mortality Ratio (under 75)</td>
<td></td>
</tr>
<tr>
<td>Standardised limiting long standing illness (under 75)</td>
<td></td>
</tr>
<tr>
<td>Proportion of pensionable age living alone</td>
<td></td>
</tr>
<tr>
<td>Proportion of dependants in single carer households</td>
<td></td>
</tr>
<tr>
<td>Proportion of economically active who are unemployed</td>
<td></td>
</tr>
<tr>
<td><strong>Psychiatric Model Variables</strong></td>
<td></td>
</tr>
<tr>
<td>Standardised Mortality Ratio (under 75)</td>
<td></td>
</tr>
<tr>
<td>Proportion of persons in lone parent families</td>
<td></td>
</tr>
<tr>
<td>Proportion born in New Commonwealth</td>
<td></td>
</tr>
<tr>
<td>Proportion of adult population permanently sick</td>
<td></td>
</tr>
<tr>
<td>Proportion of dependants with no carer</td>
<td></td>
</tr>
<tr>
<td>Proportion of pensionable age living alone</td>
<td></td>
</tr>
</tbody>
</table>

The most noteworthy features of the models were that self reported illness amongst those aged under 75 in the acute model had a strong impact, and the presence of the under 75 SMR in both models (Smith et al 1994). The proportion of elderly people living alone was also found to be a strong determinant of utilisation, and appeared in both models. The analysis considered many alternative measures of health status and social conditions. It is important to recognise that just because a specific variable is not explicitly included in the models does not mean that it is ignored. It is likely that a specific variable, to a greater or lesser extent, will be correlated with the chosen factors, and so its impact may well be accounted for in the models. So, for example, although substandard housing conditions do not appear in either model, these might be highly correlated with, for instance, the “elderly living alone” variable, and so their impact will (to the extent of that correlation) be captured in the models. In practice it can be considered unlikely that the models fail to capture any major dimension of needs.

The acute sector model was found to explain 55% of variation between small areas (Carr-Hill et al 1994a). Multilevel analysis showed that of the remaining 45% of variation that was unexplained, 20% was due to variation at the District Health Authority level. That is, almost 45% of unexplained variation in acute sector expenditure was found to be related to the policies of
local administrative bodies, and unrelated to variations in population need. The psychiatric model was found to explain 40% of the variation in utilisation between small areas. Of the remaining 60% unexplained variation, 23% was due to variation at the District Health Authority level, showing almost 40% of unexplained variation in expenditure was related to the policies of local administrative bodies.

These results suggest that expenditures are more predictable at the small area level, but this is expected as random variation within and between individuals is masked by the aggregation of the data. More importantly, these findings demonstrate the (significant) magnitude of variation that may be unexplained but is related to health scheme behaviour. Furthermore, the data did not allow estimation of provider level effects, therefore variations in provider and scheme behaviour unrelated to individuals’ needs may contribute much more than 45% to unexplained variation.

4.2 Validity of Needs Factors

Evidence on the validity of alternative needs factors in aggregate level models is relatively sparse in comparison to individual level models, reflecting difficulties in establishing the validity of population based needs factors commonly employed in capitation models (Hutchison et al 1999). These difficulties stem from the lack of available population based measures of morbidity that are independent of the supply of health services and interpretable as measures of the need for health service resources. The validity of mortality and socioeconomic measures as proxies for morbidity or health status (and hence need) therefore cannot be precisely established, since appropriate population morbidity and health status data is unavailable. However, given the paucity of population based morbidity and health status data the use of mortality rates in particular is largely inescapable, despite its potential limitations, at least until population health status data becomes available.

Evidence that has been published on needs adjustment in geographically defined populations comes largely from the UK. Statistically significant correlations (0.20-0.84) have been found between mortality and morbidity – measured in terms of self reported long term limiting illness or disability, temporarily sick and therefore unemployed, and permanently sick and not in the labour pool – at regional (Forster 1977) and sub-regional (Brennan and Clare 1978, Carr-Hill 1987) levels. Adding socioeconomic measures increases the correlation with disability (Carr-Hill 1987). National health survey data has also shown a correlation of 0.64 between SMR and self reported health status (Mays et al 1992). A correlation of 0.81 between SMR and the census based standardised illness ratio (based on long term limiting illness) was also found at the small area level (Martin et al 1995). Therefore, in the UK at least, mortality has been found to be moderately or strongly correlated with disability, long term illness and health status measures.

The Jarman index of social deprivation (a census based measure) is used to weight capitation rates for General Practice in the UK and has been found to show a positive correlation with all cause mortality for conditions amenable to GP treatment (Charlton and Lakhani 1985), and with self reported morbidity (Curtis 1990) in samples of the UK population. Hutchison et al (1997) also report that age/sex-adjusted capitation payments for GP services in Ontario did not adequately capture differences between the needs of practice populations, in comparison with analysis based on age, sex and self reported health status. The authors were, however, unable to find a...
model populated with data suitable for capitation that outperformed age and sex adjustment alone.

5  Generic Issues with Risk Adjustment Methods

A range of additional methodological issues in risk adjustment - frequently common to individual and aggregate level analysis - have recently been identified by several authors (Peacock and Segal 2000, Rice and Smith 1999, Hutchison et al 1999). These issues have received relatively limited attention in the literature to date.

5.1  Unmet Need

A largely unaddressed issue with the statistical methods used to identify and estimate the impact of needs factors, is their inability to identify and measure unmet need (Rice and Smith 1999). Statistical models seek to explain variations in actual expenditure in terms of needs factors. Actual expenditure does not capture needs of individuals that have gone unmet within the health system. Rice and Smith (1999) draw the distinction between two types of unmet need: general and specific.

General unmet need arises where services are inadequate to meet the needs of population in aggregate, for example from ‘inadequate’ global funding levels. Under such circumstances, resources within the budget constraint are most frequently allocated on the basis of relative need, as determined by the weights developed for risk adjustment. The result is no systematic discrimination between individuals covered under the scheme in terms of resource allocation and relative need.

Specific unmet need arises where specific (and identifiable) population sub-groups systematically receive fewer services to which they are ‘entitled’, than the population as a whole. Analysis to determine needs weights on the basis of historical expenditure patterns then becomes problematic, as models developed will perpetuate the existing inequity. However, even if sound methods to measure specific unmet need were available, and could be incorporated into a risk adjustment model, this would not guarantee that existing inequities would be removed. Individual health schemes under capitation may not face strong incentives to use resources allocated on the basis of need according that pattern of need. That is, without additional mechanisms to ensure the redirection of resources to those with historical unmet need, inequities may remain. Analysis of existing patterns of expenditure may therefore remain the most appropriate methods for determining needs factor weights under capitation.

Attempts to adjust for unmet need through capitation have been very limited. Rice and Smith (1999) report only one example of adjustment for specific unmet need - in the New Zealand formula for personal health care for the Maori population - who are believed to under-utilise health services. However, both Queensland and New South Wales have also made adjustments for specific unmet need in relation to the Aboriginal and Torres Strait Islander (ATSI) population. Both States have done so implicitly within vertical equity arguments – that ATSI populations should receive proportionally greater funding (over and above needs considerations) because of the
The source of this gross disparity is, in part, under-utilisation of health services by the ATSI population, and the ATSI adjustments may therefore be justified in terms of adjustments for unmet need.

The (now defunct) Queensland formula (Queensland Health 1994) suggests further research is required into the methods used in attempting to adjust for ATSI needs. The Queensland formula applied an ATSI needs index over and above other needs measures to inpatient, ambulatory and population health services components of formula. The aim of this ATSI index was to ‘to allow for the additional health services required by this population group due to their unacceptable morbidity and mortality rates’ (Queensland Health 1994). The index weighted the ATSI population in each region is by a factor of 3, which was ‘due to the fact that many health status indicators for the ATSI population are around 3 times greater than for the general population’. However, the size of the ATSI weight was not empirically justified in terms of the relationship between health status, needs, use and availability of services.

Alternatively, the ATSI weight in the Queensland formula may be justified purely in terms of a vertical equity argument – that the ATSI population should receive greater resources solely because of its historical level of disadvantage. Vertical equity considerations argue that the community as a whole may prefer more resources to be dedicated to ATSI health services, usually with the implicit notion that community preferences will reflect a general desire to reduce gross inequalities in health and health service provision. There is no reason why, in principle, vertical equity adjustments cannot be made through the use of weighted capitation formula. The approach would be - as in the Queensland formula – to attach weights to ATSI populations. These weights should reflect the community’s preferences for ATSI needs over and above preferences for the needs of other populations in society. Establishing credible estimates of these types of preferences is far from straightforward, requiring detailed population based studies using trade-off preference elicitation techniques.

5.2 Normative Utilisation

The above discussion highlights a key limitation of all empirical approaches to risk adjustment. All empirical studies necessarily examine current and past patterns of service delivery and their relationship with needs factors. Risk adjusted capitation payments are therefore based on historical expenditure patterns. Whether these expenditure patterns are appropriate or desirable cannot be answered by empirical risk adjustment studies. That is, empirical analysis can only tell us what the relationship between needs factors and past utilisation and expenditure has been, not necessarily what the relationship should be.

Different approaches are required to determine what normative utilisation patterns should be (Carr-Hill et al 1994a). Common techniques for determining normative utilisation and expenditure patterns include economic evaluation and priority setting methods. These seek to determine the optimal distribution of expenditure to achieve the maximum health related benefits for a given level of expenditure. Such methods are not typically applied at the health system level. Economic evaluations will often compare only two or three specific interventions, or just one particular
intervention at any given time, and require large scale trials to do so. This provides limited insights into health related benefits and health service needs.

However, other priority setting models/frameworks such as the QALY league table approach (see Coast 1996) or the Health sector wide disease based model (developed by Segal and Richardson 1994), are designed to consider health services utilisation and resource allocation at the health system level. Such models are potentially able to contribute to determination of the optimal size of the health budget and its efficient distribution across program types and sub-populations (Richardson and Segal 1999).

As potential health related benefits from a larger range of interventions becomes available, risk adjustment methods may be able to be extended beyond identifying the relationship between needs factors and historical expenditure. However, in the interim, risk adjustment then provides a partial approach to determining normative utilisation patterns. Even should appropriate normative adjustments be developed, the limits to the ability of risk adjustment to achieve goals in changing service mix and delivery arrangements should be carefully considered in assessing the merits of capitation within broader health system objectives. However, when used in conjunction with other policy tools, such as the procurement of cost-effective services based on evidence from economic evaluation, capitated budgets based on sound risk adjustment methods still represent a fundamental plank to meeting a range of potential health system goals.

5.3 Public-Private Mix

The treatment of the private sector in the development of public sector health service funding has represented an area of some controversy. This is particularly true in the Australian context, where the private sector accounts for over 30 per cent of all health service expenditures. The main issue is whether a public health system should consider private sector use in its funding mechanism. The answer to this issue lies in the ethical principles that underpin the health system. In a general sense, two polar principles may be adopted (Peacock and Segal 2000).

The private sector acts as a ‘pressure valve’ for the public health system. Areas or regions with relatively high private sector utilisation will use relatively fewer public sector services and resources. Those high private utilisation areas should receive less public funding, and the public resources freed up should be redirected to areas with greater demands on the public system. This standpoint defines equity in terms of equal access for equal need where access includes the perceived availability of both public and private sector services and the capacity to self-fund.

The alternative principle is that public sector should provide a comprehensive set of services for all individuals and populations, based on notions of minimum standards and core services. Areas or regions receive public sector funds relative to their underlying health care needs irrespective of the level of private sector utilisation. This standpoint defines equity in terms of equal access to public health services for equal need, and is more closely related to a rights or claims based approach to equity.

Which standpoint is the correct principle for the Australian context is a normative question. The answer would be best sought through research into community values and preferences over
health services. In the absence of such information, both Federal and some State Governments have inclined towards the first standpoint.

The Health Care Agreements which outline Federal funding policy, and some population based State mechanisms, have adjusted for private sector utilisation in funding arrangements.

Both standpoints, however, rest on the identification of the underlying needs for health care of populations irrespective of which sector meets those needs. To identify population need, a population level analysis is required which includes the need for all health services. It appears to be more prudent to take this more comprehensive approach to modelling, and to make any judgements about the treatment of the private sector after needs have been identified. In that way unambiguous indicators of the need for health services may be developed, with political and value judgements made explicit without ‘contaminating’ analytical phases of the development of a formula. This implies that analytical models should include both private sector use in utilisation measures, and the availability of private sector services as a potential determinant of variations in utilisation. The caveat to this approach is that services which the private sector provides, but the public sector does not, should be excluded where the public views those services as ‘non-essential’. A formula that allocates funds on the basis of needs for services that are not provided in the public sector, and are not considered in some way essential by the community, is undesirable. In the absence of information on what the community considers as essential and non-essential services, efforts have focussed on identifying private services that are non-substitutes for public services. In Australia, these services have been estimated to represent between 20-30 per cent of private sector expenditure (NSW Health Department 1996).

5.4 Urban and Rural Populations

Experiences with health services, and the health and socioeconomic characteristics of those living in urban and rural areas varies greatly. This implies careful consideration of the treatment of urban and rural populations in risk adjustment models is needed, as populations and individuals may display considerable heterogeneity in their health needs, access to health services and utilisation according to their geographic location.

Rural areas are frequently characterised by low population density and large distances between urban settlements. Rural communities may experience differential use of health services for several reasons (Peacock and Segal 2000). Poorer availability of specialist services may increase the time between initial presentation and confirmation of diagnosis, which may lead to increased morbidity, hospitalisation rates and/or length of stay. Similarly, worsening of patient condition after presentation to a GP may increase morbidity and risk of death if remoteness results in a delay in receiving further medical or surgical services. Differences in the need for health services and differences in their organisation and delivery will also result in utilisation patterns that are not observed for urban populations (Peacock and Segal 2000). Inpatient hospitalisation rates may be increased as GPs admit/refer more people to hospital to reduce risks of increased morbidity due to poor availability of specialised services. Outpatient and Accident
and Emergency services may be more highly used due to shortages of GPs in rural and remote areas (and in Australia a lower rate of bulk billing compared with GPs in urban areas\(^4\)).

In modelling the relationship between health service utilisation and needs, urban/rural differences suggest that the treatment of urban and rural populations requires careful consideration. A relatively simplistic approach is to define urban and rural areas by population density. This was the approach taken by the New South Wales Government in their 1993 revision of the State resource allocation formula (NSW Health Department 1993). The formula included a variable for the proportion of the population that lived in a settlement of less than 1000. But, as the authors of the study that formed the basis of the needs adjustment noted, this measure may be limited as it may produce a counter-intuitive index of rurality for some remote areas (Eckstein and Gibberd 1994). This occurs mainly where remote areas are dominated by a single settlement – for example a mining town – but where the whole of the area is geographically very isolated and has poor access to health services.

The most recent revision of the New South Wales formula has addressed many of the deficiencies of more simplistic approaches to modelling rural and urban differences for health service resource allocation. The regression analysis of utilisation and needs now includes a more sophisticated index of population location, called the Health Related Rural Status index (Eckstein and Gibberd 1994). The new index was developed by taking account of four prime considerations. The index should be independent of socioeconomic status measures and reflect only characteristics relating to location, as the relationship between rural deprivation and conventional socioeconomic measures of depravation (income, occupation, housing and education) is poorly understood. Population density is not a robust measure to be included in an urban/rural index, as it takes no account of the distribution of population within areas used in regression analysis. Interval properties of an urban/rural index are questionable over the whole distribution, making the use of an ordinal scale a more appropriate strategy. And, the index should be specific to health, reflecting the distance to major referral centres rather than commercial centres. These considerations were used to define remote, rural, major urban and metropolitan areas classified by distance to referral centre. These classifications were then combined with data on land use to construct an ordinal index with 17 categories for urban and rural areas.

This approach represents a major advance on previous attempts to address rurality and health service resource allocation in Australia. A major remaining task is to attempt to more robustly estimate appropriate measures of deprivation in rural areas, where conventional economic and public health indicators may have only limited value. As the architects of the New South Wales formula state ‘hardship associated with limited access to physical, cultural and social amenities causes complex changes to behavioural patterns when compared to urban dwellers’ (Eckstein and Gibberd 1994).

However, the use of population density in urban/rural indices should be explored more fully. Its exclusion from the New South Wales approach was on statistical grounds: that intra-area variation in population density would make population density an inappropriate index of urban/rural populations at the small area level. This approach is sensible with small area data, but population density may still be an important component of an urban/rural index at other levels.

\(^4\) The presence of an up front fee (albeit largely reimbursable) in the absence of bulk billing will encourage greater use of hospital based services which are free at the point of consumption.
of analysis. Research should be developed using individual level data and the effects of population density on health service utilisation, including analysis of hierarchical effects through techniques such as multilevel modelling.

Analysis at the individual level would avoid the problems encountered with small area analysis, and allow greater use of a larger range of measures of urban and rural areas.

5.5 Cross-Boundary Flows

Where health service funding is based on a capitation payments, a common issue for individual health schemes is how to deal with “cross-boundary flows”. Cross-boundary flows typically occur in two ways – where a health scheme allows utilisation by patients who are not enrolled with that health scheme, and when health scheme enrollees can utilise health services not provided by that scheme (Hutchison et al 1999). Under some health system models cross-boundary flows are expressly prohibited, but this is relatively uncommon. For instance, in most health systems patients who are travelling away from home may occasionally have to make use of health services provided through another health scheme. Similarly, services provided by a given health scheme may not be comprehensive, and enrollees may make use of other health facilities where specific services are not covered by their scheme. In both situations a transfer of funds may be required from the scheme receiving a given individual’s capitated payment, to the scheme providing services. In general, the problem of the treatment of cross boundary flows becomes greater when geographically defined populations are small and/or the range of services provided within a given region is limited.

Approaches to making this transfer vary. The Canadian funding model operates a charge back or negation system where the provider with which a patient is enrolled is charged for any services a patient receives elsewhere (Hutchison et al 1999). The rate of charge back varies between 50-100% of the fee-for-service costs. The system has drawn some criticism due to its administrative burden, and the lack of control that physicians perceive they have on outside use. Response to these criticisms has included: the exclusion of “appropriate” outside use (eg emergency treatment whilst travelling, use of specialist services) from charge back; the use of a pre-determined percentage adjustment to capitation rather than monthly charges, based on historical outside use patterns; and, increased access to outside providers patient treatment records for health schemes. New South Wales adjusts capitation allocations to areas on the basis of estimated flows – unlike Canada individual cases are not identified and costed. In the UK the problem of cross boundary flows was largely eliminated through the introduction of the purchaser-provider model. Health Authorities, and now Primary Care Groups, are allocated funds on the basis of population irrespective of where they procure health services for that population.
6 Supply Side Factors and Risk Adjustment

Given the vast array of alternative health system models in the world, the approach to risk adjustment adopted by different countries has been remarkably similar. The primary goal of both individual and aggregate level risk adjustment methods has been the prediction of the expected costs of health services for individuals within health schemes. Both individual and aggregate level methods have used regression analysis, and methodologically are extremely similar.

It is useful, however, to draw a distinction between broad approaches to risk adjusted capitation under two types of health systems. The two health system models are best described as the competitive health scheme model and the non-competitive health scheme model, which have given rise to some differences in approaches to risk adjustment and the determination of capitation rates (Rice and Smith 1999). These differences primarily relate to the inclusion or exclusion of some supply side factors from risk adjusted capitation funding models.

6.1 Competitive Health Scheme Models

Under the competitive health scheme model health schemes are allowed to compete for enrollees under a range of possible financing and delivery arrangements, for example competitive health insurance, managed competition etc. Capitation schemes under this model place what Rice and Smith (1999) term a “bounty” on the head of potential health scheme enrollees, and health schemes are allowed to compete for enrollees in order to receive this bounty. Typically, this bounty is the focus of risk adjustment in competitive health scheme models. Three common features of risk adjusted capitation payments (bounties) in competitive models are (Van de Ven and Ellis 2000):

- dependent only on individuals relevant risk characteristics (risk adjustment)
- worth a precisely specified amount of money per individual
- independent of the premium paid by the individual or the specific insurance scheme chosen

If health schemes are allowed to freely set risk adjusted premiums for enrollees under the competitive model, the range of needs factors used to set premiums and the range of premiums for enrollees will become large. Under unregulated premiums, the average premium for an enrollee has been estimated to vary by a factor of 10 for the same level of health coverage (Van de Ven et al 1997). Equity concerns over premium rates in most countries, particularly for the sick/high risk, have led to the regulation of health schemes to reduce variation in premium rates. This is commonly achieved through community rating insurance premiums over a range of factors, which create implicit subsidies from low risk to high-risk enrollees within a given scheme (implicit in the sense that individual's premiums do not vary).

However, this creates potentially predictable losses and profits for health schemes for low and high-risk enrollees respectively. Similarly, imperfect predictors used in any risk adjusted payment from funders may still result in predictable losses and profits from different risk enrollees. This creates the potential for selection – actions by consumers and health schemes to exploit unpriced risk heterogeneity (Newhouse 1996).
Selection occurs in two forms – adverse selection and cream skimming. Adverse selection refers to the incentive for high risk individuals to buy more coverage than low risk individuals within the same premium risk group. This occurs when individuals know when they are high or low risk, but health schemes do not (asymmetric information). This may result in an unstable insurance market where low-risk individuals will try to separate themselves from the high risk by seeking other coverage packages, and the premiums for those high risk remaining in the original insurance scheme will spiral.

Cream skimming refers to health schemes preferences for low risk enrollees over high risk enrollees within the same payment risk group. This also occurs under asymmetric information, but only when health schemes know which individuals are high and low risk, and the individuals themselves do not know. Health schemes may cream skim in a number of ways. If schemes can observe, ex ante, high risk individuals (within a given premium risk group) they may structure insurance schemes to be unattractive to those high risk individuals (eg high deductibles, exclusion of drugs etc). Health schemes may also, ex post, design coverage to deter the high risk (eg by limiting access to the “best” specialists, limiting supplementary insurance options etc). Schemes may furthermore selectively target low risk individual’s in advertising, or offer high risk enrollees “payoffs” to terminate insurance coverage.

The principle role of capitation in this type of model is the promotion of efficiency in the operation of the insurance market (Rice and Smith 1999). The main aim of risk adjustment is therefore to minimise the potential for adverse selection and cream skimming by reducing variation in the difference between reimbursement and actual/expected expenditure. As a result, the main aim of risk adjustment under the competitive health scheme model is to ensure that capitation payments are adequate to cover predictable future expenditures on insured health services provided to enrollees (Hutchison et al 1999).

### 6.2 Non-Competitive Health Scheme Models

Under the non-competitive health scheme model health schemes are not permitted to compete for enrollees, for example enrolled populations are determined by geographically defined administrative regions, employee insurance schemes etc. Capitation schemes under non-competitive health scheme models have typically arisen where the responsibility for purchasing health services for a given population have been retained within the public sector.

Closed (or fixed) enrollment to health schemes under the non-competitive model removes explicit selection problems. High risk individuals may seek greater coverage within the same premium risk group (ie the entire enrolled population), but closed enrollment means that low risk individuals cannot then seek coverage elsewhere. Moreover, contributions from individuals are not likely to vary between health schemes under social insurance systems (which are common in the non-competitive model). Similarly, opportunities for health schemes to explicitly cream skim are removed if health schemes are obliged to treat a defined population. However, if health scheme expenditure on high and low risk patients and quality of care is not directly observable, some opportunities for “scrimping” may still remain.
In the vast majority of countries with a non-competitive health scheme model, the major role of capitation is the pursuit of explicit equity objectives (Rice and Smith 1999). Questions concerning relative funding levels and relative population need are inevitably linked to equity issues in the provision of health care, whether equity is considered implicitly or explicitly in funding arrangements. In health services, equity typically relates to the distribution of some factor in relation to the health care needs of individuals and populations. Hutchison et al (1999) therefore suggest the purpose of capitation under this health system model is to ensure funding is consistent with the relative needs for services of enrolled populations. Typically, explicit objectives in capitation under a non-competitive health scheme models relate to equity in terms of equal funding for equal need or equal government subsidy for equal need across health schemes which individuals are compelled to join (Rice and Smith 1999).

Note however, that the predominance of equity objectives in capitation funding under existing non-competitive health scheme models does not preclude the potential pursuit of other objectives. Capitation funding in a non-competitive model may be used solely to pursue efficiency objectives, or a combination of efficiency and equity objectives. For example, capitation in the context of the Australian Coordinated Care Trials may seek solely to pursue efficiency goals in terms of cost-effectiveness. This would be achieved through Funds Pooling to remove barriers between health sector budgets to promote (cost-effective) service substitution. The focus on equity in non-competitive models reflects the objectives of the health systems where capitation funding has been used in recent years. These objectives may not apply in other contexts.

The predominance of equity objectives as the main purpose of capitation under the non-competitive health scheme model may be seen as a combination both of broader health system objectives and the organisational characteristics of health systems. In general, capitation under this health system model has been used to allocate funds from a central funding agency to devolved health schemes. The absolute level of the budget to be shared between health funds has been, in the vast majority of countries, exogenously determined. Capitation has then typically sought to distribute these funds in a manner consistent with some underlying equity objective(s). A range of other potential goals, such as cost containment and reducing the potential for cream skimming, then become largely irrelevant, as the global budget is capped and most frequently, health schemes cannot cream skim. To the extent that allocating funds on the basis of need will reflect capacity to benefit from health services, allocative efficiency is a common secondary goal of capitation under this model. However, allocative and technical efficiency goals are more usually addressed through other tools at the health scheme level, such as cost-effective purchasing strategies.

### 6.3 A Synthesised Approach to Capitation

All of the potential objectives for capitation discussed above will be appropriate in the majority of health systems, but with varying degrees of importance. For instance, under a non-competitive model with geographically defined populations, cream skimming may not be possible, but there may be some risk associated with the financial viability of health schemes. Despite differences in

---

5 Note, this is only a hypothetical example of a potential objective of capitation in the Coordinated Care Trials. Equally, equity may be an important objective, evidenced by the guarantee that individuals enrolled in the Trials retain full access to Medicare and Pharmaceutical Benefits entitlements.
the objectives of capitation in alternative contexts, and the relative importance attached to
different objectives, the common theme of risk adjustment is that it seeks to “best explain”
variations in expected costs over a fixed period of time.

This allows the prediction of the future health service costs for individual’s which form the basis
for prospective payments to health schemes. However, what is meant by “best explain” may vary
subtly between different types of health system models and under alternative objectives for
capitation, and may imply different sets of weights in any capitation formula. Specifically, the
treatment of needs and supply factors in explaining variations in expected costs has differed
between the two predominant approaches to capitation - risk adjustment to promote efficiency in
the insurance market in the competitive model, and risk adjustment to promote equity in the non-
competitive model. These differences in the treatment of needs and supply factors provide the
main distinction between approaches to capitation.

Capitation to promote efficiency in the insurance market in the competitive model has focussed
on explaining expected cost in terms of both needs and supply factors. Capitation payments have
then been based on predicted costs that incorporate both individuals’ health service needs and
supply side factors as sources of variation in costs. As a result, capitation payments under this
approach have not generally sought to correct, or adjust, for variations in local clinical policy, the
managerial efficiency of providers, and variations in the local availability of health services.

The reason for this approach lies in the objective of promoting efficiency in the insurance market.
Attempts to correct for variations in managerial efficiency under capitation, for instance, may
promote cream skimming behaviour. If a relatively inefficient (high cost) provider is reimbursed at
the average efficiency rate of payment then it will seek to reduce costs (improve efficiency). One
potential means of reducing costs is to cream skim. Therefore, attempts to influence supply side
behaviour through capitation payments may create perverse incentives that run contrary to the
underlying goal of capitation in a competitive health scheme model. This has led to a focus on
capitation payments that reflect variation in costs in competitive health scheme models,
irrespective of the source - need or supply side - of that variation.

Capitation to promote equity in the non-competitive model has focussed more closely on
explaining expected cost in terms of needs factors. Capitation payments have most commonly
been based on predicted costs that include individuals’ health service needs as a source of
variation in costs, but which correct, or adjust, for supply side sources of variation in costs. The
correction for supply side sources of variation in costs implies each health scheme receives
payment based on an individual’s health service needs given some standard level of
circumstances.

In general, reimbursement occurs at a rate which reflects the expected cost for an individual with
a given level of need adjusted for the average impact of variations in clinical policy, managerial
efficiency, and the availability of health services on costs. That is, supply side variations in cost
are smoothed, or averaged, across health schemes. The effect is that a health scheme that has
higher utilisation, and hence costs, because of historically high access to services receives the
same payment as a scheme with low utilisation due to historically poor access. In this way,
capitation may be used to correct existing inequities in service provision between health
schemes, by using relative resource shifts from high access to low access areas. Similarly,
reimbursement rates which adjust for supply side influences may be used to reduce local practice variations and managerial inefficiency.

If supply side sources of variation in costs are not adjusted for, an equity goal of equal resources or equal access for equal need may be undermined. If a health scheme with historically higher access levels, and therefore higher utilisation and costs, is rewarded on the basis of its historically high costs then capitation may perpetuate and promote existing inequities in access. This has led to a significant focus on capitation payments that reflect needs-based variation in costs, but which correct for supply side influences in non-competitive health scheme models.

The above discussion is complicated somewhat when the exact nature of supply factors in considered more fully. Supply factors, as outlined above, relate to variations in the efficiency of health service providers and variations in their managerial and clinical policies. Some aspects of efficiency and policy will be under the control of service providers and unrelated to the health service needs of individuals covered by schemes. Supply side variations in costs that may be directly influenced by health schemes and that are unrelated to need may be termed avoidable cost differences. It is these factors which non-competitive capitation models that have both equity and efficiency objectives have sought to correct for in capitation payments.

However, some aspects of efficiency and policy will lie outside of the direct control of service providers. These supply side variations in costs may be termed unavoidable cost differences. Unavoidable cost differences are generally due to variations in local factor input market conditions and in location-specific costs of service provision. Local factor input market conditions may affect provider costs where local wages and rents vary according to the extent of local labour market conditions, availability of land etc. For example, wages in remote areas may be higher than in urban areas, reflecting local labour market conditions and the need to offer higher wages to attract staff to health services in remote locations. Similarly, providers based centrally in major cities may face higher land rents than those based in more rural locations. Location-specific costs of service provision may affect provider costs where the physical location of health services and populations influence the costs of service delivery. For example, service delivery to remote and sparse populations may be inherently more costly than in urban areas due to higher transportation costs. Under both competitive and non-competitive models capitation has generally sought to cover differences in health scheme costs where they are deemed to be unavoidable.
6.4 Supply Side Factors in Empirical Studies

(i) Supply Side Factors in the US and The Netherlands

The US Medicare AAPCC formula fully adjusts for all variations in local area relative health care costs using county level average indemnity payments by age and sex groups. These indemnity payments are intended primarily to reflect input price variations – unavoidable cost differences that are exogenously determined.

However, whilst indemnity payments may be a reasonable measure of variations in input prices (Van de Ven and Ellis 2000), the AAPCC formula assumes that indemnity payments are only minimally affected by variations in health service supply, provider practice patterns, management practices and the care-seeking behaviour of beneficiaries (Hutchison et al 1999). This assumption is highly questionable, and as indemnity payments will almost certainly reflect practice style variation and other management practices. AAPCC capitation payments therefore almost certainly overstate the variation that policy makers should use in calculating risk-adjusted subsidies (Van de Ven and Ellis 2000), and may perpetuate existing inefficiencies and inequities in service provision.

Similarly, the urbanisation and service type adjustments used in the Dutch risk adjustment model do not attempt to separate out avoidable and unavoidable cost differences. All sources of cost variations between providers are therefore reimbursed for, under an implicit assumption that health schemes cannot control these provider cost variations.

(ii) Supply Side Factors in the UK

Rice and Smith (1999) note that whether a factor is considered ‘legitimate’ may be a matter of the policy context within which the capitation scheme operates. This applies, in particular, to the use of a range of factors which may be added to capitation weighting over and above needs weights based on individual and population characteristics. Typically such additional adjustments have related to ‘unavoidable’ differences in the costs of providing services between health schemes. Examples of such unavoidable cost differences include adjustment for local labour market effects on wages in England and for differences in the costs of service provision between urban and rural/remote areas in New South Wales.

The English approach has been to assume health schemes cannot influence input prices in general, and capitation rates are adjusted using general wage and land price measures. They do not use health service wage rates to avoid the potential for health schemes to influence local rates and employment practices to influence future capitation rates. Adjustment for unavoidable geographic variations in the cost of providing health services through the Market Forces Factor (National Health Service 1997). This factor adjusts for variations in the costs of labour, land, buildings and equipment. The major element of the market forces factor is the labour weight, based on regression analysis of the impact of location on labour wage rates outside the National Health Service (Wilson et al 1996). This approach assumes that external wage rate variations are reflected proportionately in the variation in unavoidable costs incurred by NHS providers, and significantly reduces the potential for perpetuating inefficiencies by excluding provider pay structures.
The York formula methods, however, highlight a particular methodological problem, almost completely ignored by other risk adjustment models - the correct treatment of supply in the model of the demand for health services. Due to the doctor-patient agency relationship the supply of and demand for health care cannot be separated as in the orthodox economic model of supply and demand. In health services, the supply of health care may affect demand through two mechanisms.

First, where there is excess demand, supply may constrain utilisation - which is consistent with the orthodox economic model. However, market forces in the orthodox model mean that supply will (eventually) expand, and excess demand will be removed.

In the context of health services, markets are largely absent, or where present, operate extremely inefficiently. Therefore, unmet demand is likely to be present in all health systems. Furthermore, the supply of health services often reflects both the needs of the population (ie unmet and expected demand) and a range of factors unrelated to the need for services. Examples of factors unrelated to the need for services include political decisions over hospital location, historical concentrations of teaching hospitals in capital cities, lifestyle decisions of GPs and the desirability of alternative residential locations etc.

Second, the imperfect agency relationship leads to the possibility of supplier-induced demand (SID). It is possible under imperfect agency that suppliers may influence the level and type of services provided, because individual’s do not have complete, perfect information. Therefore, unlike in the orthodox model, suppliers may be able to induce demand to satisfy their goals, and this may produce patterns of utilisation that do not fully reflect underlying population need.

This framework was central to the two key advances in the York risk adjustment methods – to statistically isolate and estimate the impact of needs factors on utilisation independent of the influences of supply side factors, and to adjust for policy effects at the health department level. This was achieved through the estimation of the direct and indirect impact of needs factors on utilisation, and the use of multilevel modelling techniques, as outlined above.

The implication for risk adjustment methods is that the inclusion of supply side factors in a risk adjustment model is necessary to develop a robust and valid predictive model of health service utilisation and expenditure. Omission of supply side factors will result in a loss of predictive power and potentially introduce bias into the regression model. Supply measures are in part determined by population needs, therefore appropriate statistical models that “recognise” simultaneity in the model should be used. However, directly including all supply side factors in the weights used to determine capitation payments may produce inappropriate incentives (ie reward inefficiencies and inequities). That is, if observed variation in costs is explained by supply side factors that are attributable to historical or residential features then the inclusion of these supply side factors will perpetuate these undesirable determinants of utilisation.

Risk adjustment models should therefore seek to determine what part of supply is determined by needs, and what part is determined by non-need related factors. Examples of non-need related factors include historical and political decisions on the location of health services, local clinical and managerial practices, and health department or state government level policies. Capitation payments should then be based only on adjustments made for needs factors and supply factors.
outside the control of health schemes. This requires both careful treatment of supply side factors in regression analysis, and the use of more sophisticated modelling techniques – such as multilevel modelling – which “purge” the model of the effects of local variations in practices.

The correct treatment of needs and supply factors has raised some concern across the countries studied by Rice and Smith. For instance, in Belgium debate and concern over supplier-induced demand has led to the exclusion of physician supply from regression equations – ie health schemes are not compensated for variations in physician supply available to their beneficiaries, even though the schemes may have no control over the consequent variation in utilisation as they are not allowed to negotiate service mix with providers.

However, in the US Medicare formula variations in local expenditure are included in payment adjustments, even though supply may be an element of this effect. With the exception of the York approach, the treatment of supply has therefore been, at best, very partial. Rice and Smith (1999) suggest the separation of legitimate needs and illegitimate supply effects on variations is a remaining major problem in determining capitation rates. Many countries are promoting and perpetuating both inequities and inefficiencies by rewarding inefficient behaviour, and providing greater funds for individuals and populations with already high levels of access to health services.

7 Risk Adjustment for Capitation Funding

The preceding discussion has focussed almost exclusively on the validity of approaches to risk adjustment and their performance in terms of predicting health service costs. However, even if the best available risk adjustment methods are used a range of other issues require consideration if predicted costs are to be used to determine prospective budgets that provide incentives consistent with the goals of the health system. These issues relate to the translation of predicted costs from risk adjustment methods into a risk adjusted capitation funding models. Specific issues relate to the population and services covered by health schemes in the funding model, the ability of schemes to risk pool, risk sharing arrangements, and risk management in the implementation of capitated budgets.

7.1 Budget Setting and Scheme Coverage

The process of developing a capitation model can be broken down into three major decisions (Rice and Smith 1999):

(i) the amount of finance to be distributed for the service in question
(ii) the factors to be incorporated into the capitation
(iii) the weights to be placed on those factors

The first decision involves addressing the interrelated issues of the range of services and health schemes to be covered by the capitation scheme and the population(s) that are expected to be covered by health schemes. These issues may be grouped under the heading of capitation scheme coverage in terms of budget, services and population, and are addressed below.
The second and third decisions relate to approaches to risk adjustment: the methods and processes used to determine the influence of individuals’ characteristics on health service needs and to incorporate those influences into capitation payments. These decisions will be, in part, determined by decisions relating to capitation scheme coverage, but have received much greater attention in the literature. The factors incorporated into capitation models, and methods for determining appropriate weights were considered in sections 3 to 6.

(i) Service Coverage

In the majority of capitation schemes, service coverage has been determined by existing organisational boundaries and corresponding responsibilities. These responsibilities may encompass a relatively narrow range of services for a specific sub-population, or a wide range of services for an entire state or national population including primary, secondary and tertiary care and welfare services. Typically, capitation has been employed to address specific funding and delivery objectives within a given organisation, with a given budget, based on existing organisational arrangements. Service coverage has therefore received relatively little attention in the literature, as most capitation models have taken the range of services to be covered as fixed and exogenously determined.

However, wider service coverage is generally more desirable than narrow service coverage under capitation funding. The greater the comprehensiveness in services covered, the greater is the potential for flexibility in resource allocation and therefore the greater are the incentives for allocative efficiency and cost-effective use of resources. For example, capitated budgets which cover primary, secondary and tertiary care create greater incentives for allocative efficiency than budgets covering only secondary care. They also reduce incentives for cost shifting across budget boundaries. Introducing artificial boundaries through budget rules impedes substitution between services and thus the achievement of allocative efficiency.

(ii) Size of the Global Budget

Under the non-competitive model capitation schemes have frequently been used to allocate funds from a defined budget. For example, the UK capitation models for hospital and community services allocate funds from a defined budget which is exogenously determined through negotiations between the Health Department and the Treasury. As a result, the capitation literature from countries using non-competitive models has tended to focus on methods of risk adjustment rather than determination of the overall budget to be allocated.

The key reason why the budget may be determined exogenously under the non-competitive model is that, historically, health schemes have not been allowed to compete for enrollees in the majority of such systems (e.g. populations under health schemes are defined by geographic areas). Capitation with no competition for enrollees is then solely a relative concept. The overall size of the global budget for allocation will be distributed according to the relative weights of any capitation model to reflect relative differences in expected costs. Whether the resulting distribution of funds matches the absolute expected costs for any given individual will also depend on the size of the global budget. There is no guarantee that a sound risk adjustment mechanism will result in the expected costs of individuals being fully covered for health schemes. If the overall budget is exogenously set at an “unrealistically” low level, then capitation payments will be lower than expected cost.
Under a competitive health scheme model, health schemes are allowed to compete for patients. Under this approach the global budget may not be predetermined. For instance, a health insurance company may not set a fixed budget each year, rather the sum of the capitation payments made to its health schemes may make up the (retrospective) budget total. Total capitation payments will reflect an absolute dollar value per enrollee, determined by risk adjustment, and the number of enrollees in health schemes. Capitation models may then have a key role in determining the global budget. Government budgets may be uncapped resulting in a retrospective budget total derived from total actual expenditure. Under a capitation approach the “budget” would be set in an identical manner to under private health insurance: the budget would represent the sum of capitation payments made to health schemes, reflecting payment rates and number of enrollees.

(iii) Population Coverage

Population coverage often presents a non-trivial problem for determining capitation rates. Estimation of the size of the population pool in each scheme is of fundamental importance. In non-competitive systems where populations are geographically defined this task may be relatively straightforward. Census and population data can be used to provide relatively robust population estimates, even where census small area boundaries do not coincide exactly with health scheme administrative regions.

It may also be difficult to estimate population for competitive health insurance schemes. Whilst point estimates may be very accurate for these schemes, based on number of enrollees, it may be much more difficult to predict future demand and hence future number of enrollees. A range of factors may affect health insurance enrollee rates at any time, many of which may not be predictable or within the control of the health funds (eg introduction of tax rebates, downturn in economic expectations etc). Enrollment estimates may be most fragile for some forms of public health funds. For example, allocations to British NHS GP fundholders were made on the basis of GP list size (patients have to register with a single practice), but list size estimates are generally held to be unreliable. The problem is compounded where patients do not enroll with a specific scheme and are only “registered” when they use service (for example GP services in Australia), or when they change scheme frequently. In these situations the degree of uncertainty as to the health scheme population and their characteristics may even preclude the use of capitation based funding.

7.2 Financial Risk and Risk Pooling

It is important to recognise that whilst capitation notionally assigns a dollar amount to each individual, there may be substantial variation in realised costs at the health scheme level. There are two main sources of this variation. Firstly, risk adjustment can only partially predict future health expenditures. Under individual level analysis up to 75-80% of variation in costs may not be predictable irrespective of needs factors used in modelling. Secondly, there is no obligation that health funds should spend the full capitated amount on each enrollee. Health schemes and providers typically have a significant level of discretion in the services provided to any given individual, and may seek to cross subsidise relatively expensive treatment for some patients from
relatively inexpensive treatment of other patients. Furthermore, health schemes may have little control over the services provided by its providers, which may be source of significant variation in costs.

Health schemes therefore have to manage the risk associated with variations in realised cost, whether from imperfect risk adjustment or from local variation in policies and practices. Under a non-competitive model failure to manage risk and risk pool appropriately may result in increased waiting lists, poorer quality of services, refusal of treatment, and/or financial insolvency for health schemes. Under a competitive model these responses may also occur, but schemes in addition face significant incentives for cream skimming in order to remain profitable. Given the poor predictive power of risk adjustment models, and the extent of discretionary practices, risk management and risk pooling under capitation has been the focus of increasing debate in recent years.

The relatively poor predictive power of risk adjustment models has implications for risk pooling and the size of the population covered by each scheme. Imprecision in risk adjustment essentially places lower limits on the size of the population to which budgets should be allocated. As imprecision - unexplained variation in the model – rises, financial risk to the health scheme may increase.

Newhouse et al (1989) outline a model of individual level health expenditure, where expenditure is determined by three main factors:

\[ \text{Expenditure} = bX_i + u_i + e_{it} \]

where \( i \) is the individual, \( b \) is the population average response to a vector of known and measurable needs factors \( X \), \( u \) is a time-invariant component of the variance associated with individual \( i \), and \( e \) is a time-variant component of the variance associated with individual \( i \). The component \( bX \) therefore represents the variation which is predictable in the model, \( e_{it} \) pure random variation, and \( u_i \) variation which is due to unobservable (and therefore omitted) factors, such as unobservable patient characteristics and clinical practices. Estimation of this model therefore consists of parameter estimates for population average responses to needs factors \( b \) and for the residuals \( e_{it} \) which capture both \( e_{it} \) and \( u_i \).

The residuals from risk adjustment regressions, \( e_{it} \) are most frequently assumed to be normally independently distributed with a mean of zero and standard deviation of \( \sigma \). They represent the difference between actual expenditure and expenditure predicted by the risk adjustment model. A key issue for each health scheme is the variance associated with these residuals at the scheme level, which turns out to be closely related to the number of enrollees it takes on.

Residuals from risk adjustment models are based on analysis of a total population, \( N \), from which each health scheme enrolls a sample of patients \( n \). From statistics theory, the central limit theorem shows that even for small values of \( n \) (eg 20-30), the sampling distribution of residuals for health scheme enrollees will be normal. The sampling distribution mean value will also be the population mean, which in this case is zero. Of primary importance for health schemes is the variance of this sampling distribution. The standard error of the sampling distribution (from the normal approximation rule) is given by:
Therefore, as the sample size – the number of health scheme enrollees – increases the variance of the sampling distribution falls, producing a more peaked distribution centred on a mean of zero. This is the essence of risk pooling – by increasing its sample of enrollees a health scheme can reduce variability in the difference between predicted and actual expenditure.

The two key variables for determining financial risk are therefore the variance in the risk adjustment model residuals, and the sample size each scheme enrolls as patients. Unfortunately, published articles do not routinely publish information relating to variance in the residuals from individual level risk adjustment models. This is a significant gap in the literature. Research is urgently required that uses local individual level data to analyse of unexplained variation in residuals and their implications for the financial viability of budget holding health schemes.

However, Van de Ven and Ellis (2000) have suggested that financial risks to schemes become “small” when enrollee numbers reach approximately 5000 and when risk adjustment is based on individual level data. This suggestion is not backed up with the empirical analysis used to arrive at this figure, however, so some caution may be warranted.

The only study that has directly explored this issue examined risk to GP budget holders from funding allocations made from the UK aggregate level risk adjustment model (Martin et al 1998). The aggregate level analysis was undertaken using small areas, with an average population size of approximately 10,000. This is also approximately the average number of enrollees in a GP fundholding practice. The standard deviation from risk adjustment modelling was 0.1062 (using a logarithmic model). Martin et al then demonstrated the impact of the use of the aggregate level results for financial risk for GP fundholders in two ways.

Firstly, 95% confidence intervals were estimated for fundholders – ie the limits within which actual expenditures will fall in 95% of health schemes. They found that where health schemes had n=10,000 enrollees (the actual size of many fundholding practices) 95% of schemes fell within approximately ±20% of expenditure predicted under risk adjustment. This implies a need for significant tolerance for budget under and overspends. However, increasing scheme size reduced the magnitude of these deviations, falling to ±7% when scheme size was increased to n=100,000. Secondly, the probability that a scheme would be ±10% from predicted expenditure was estimated. Results showed that for schemes of n=10,000, 1 in 3 would experience actual expenditure that deviated by more than ±10% from predicted expenditure, whereas for schemes of n=100,000 the probability was dramatically reduced to 1 in 400.

These results demonstrate the significant level of financial risk an individual health scheme may face. Straight application of the English capitation model to GP fundholders could result in financial insolvency for the majority of schemes if counterbalancing measures were not employed (such as those discussed below). It is mainly for these reasons that the UK has struggled to develop a risk adjusted capitation model for allocating funds to GP fundholding practices. The results also demonstrate the ability of schemes with large enrollee bases to risk pool. Using UK data, a ten fold increase in scheme size suggests over a hundred fold reduction in financial risk. Most importantly, the study by Martin et al demonstrates the limitations of aggregate level analysis. Risk for schemes is very high when enrollee numbers are small, primarily because the unit of analysis (small areas) is large. Under individual level analysis the risk to health schemes...
with small numbers of enrollees is reduced because of increased precision in the estimates of individual enrollee’s health service costs.

7.3 Ex Post Risk Sharing Arrangements

A number of ex post risk sharing arrangements may be used to counter the potentially adverse consequences of risk for health schemes (insolvency, cream skimming etc). Ex post risk sharing usually implies that some health scheme costs are retrospectively reimbursed by the funder. This reduces the potential for scheme insolvency and incentives for cream skimming, but the use of retrospective reimbursement also dampens incentives for efficiency (Newhouse 1996).

The goal of risk sharing is not, however, the reduction of health scheme’s financial risk through reducing variation in expenditures (which is more commonly achieved through reinsurance). The goal is to reduce health scheme’s predictable profits and losses, whilst preserving their incentives for efficiency in as far as is possible (Van de Ven and Ellis 2000).

Ex post risk sharing arrangements may take a number of alternative forms - cost-based risk sharing, condition-specific risk sharing, service-specific risk sharing.

(i) Cost-Based Risk Sharing

Cost-based risk sharing strategies typically take one of three forms – outlier risk sharing, proportional risk sharing, and risk sharing for high risks.

Outlier risk sharing (or stop loss) arrangements take the form of retrospective reimbursement by the funder for individuals who reach a critical (or catastrophic) threshold level of health expenditure for any given period. This sets a ceiling on the total level of expenditure a health scheme is liable for, for any given patient. Evidence on outlier risk sharing suggests it may offer some protection against the risk of unexpectedly high expenditures at a relatively modest cost (Beebe 1992). Using 1992 US Medicare data, and assumptions of outlier retrospective reimbursement at 45% for critical thresholds between US$10,000-US$100,000, the percentage of enrollees exceeding the threshold was found to lie between 0.07-11.1% with retrospective outlier reimbursement making up 0.14-19.5% of total expenditures (Beebe 1992). Keeler et al (1998) simulated the effects of retrospective outlier reimbursement set at 10% of total expenditure, and found it reduced under and over-payments (under capitation) by 35-50%.

Proportional risk sharing (or cost sharing) arrangements typically consist of retrospective reimbursement by the funder for a proportion of all health expenditures incurred by a scheme. The funder therefore shares financial responsibility for all individuals with each health scheme. Proportional risk sharing has also been found to reduce under and over-payments under capitation by 35-50%, when the retrospective reimbursement rate is set at 25% of scheme expenditure (Keeler et al 1998).

Both these approaches to risk sharing do not require the funder, or health schemes, to identify ex ante predictable losses. It is those predictable losses that provide incentives for cream skimming. They are instead designed primarily to protect health schemes against ex ante unpredictable
losses (ie high expenditures – eg previously healthy car accident victims), thereby reducing the risk of schemes becoming insolvent. The cost of reducing the risk of insolvency is dampened incentives for efficiency.

Risk sharing for high risk arrangements have primarily sought to reduce incentives for cream skimming. To reduce incentives for selection Van de Ven and Van Vliet (1992) have proposed that health schemes could specify a group, or percentage, of enrollees prior to the budget period, over which it wishes to share risk with the funder. This would place the onus on health schemes to identify high risk patients ex ante and therefore reduce its predictable losses. This would serve to reduce incentives for selection against high risk individuals, where risk is not fully captured in premiums.

Risk sharing for these high risk enrollees may apply to a percentage of all expenditures, or to expenditures over a critical threshold level, or a combination of both. This approach may be particularly useful for small groups of enrollees with predictable high losses, ie where risk adjustment has not adequately captured the determinants of very high cost patients. Incentives for diagnosis up-coding to attract extra retrospective funding are also removed. In addition, the (often asymmetric) information held by the scheme regarding high loss patients may be passed onto the funder under these arrangements, which would allow further research and refinement of risk adjustment for very high cost patients. However, ex ante unpredictable losses would not be covered under this approach, leaving greater potential for health schemes to become insolvent.

Evidence on the use of risk sharing for high risks has found it to be the most effective risk sharing strategy for reducing cream skimming. Estimates suggest retrospective reimbursement for enrollees in the highest 4% for prior year expenditures would reduce predictable losses to health schemes by between one half and two thirds (Van Barneveld et al 1996, Van Barneveld et al 1998). These studies did not examine whether the risk of insololvency changed as a result of risk sharing for high risks.

(ii) **Condition-Specific Risk Sharing**

An alternative approach to risk sharing is to retrospectively reimburse health schemes for the occurrence of specific medical conditions from prospectively determined payment schedule (Enthoven 1988). This approach is best suited to conditions where diagnosis is difficult to manipulate and treatment is high cost and relatively non-discretionary. Prospective payment rates (such as under case payment rates) have the advantage that they do not dampen incentives for efficiency. Incentives for cream skimming will be reduced to the extent that a health scheme can ex ante predict above average expenditure for the specified conditions within its enrolled populations. This approach has the advantage that health schemes specialising in the treatment of high risk patients will receive appropriate funding for each high risk individual. Condition-specific risk sharing has been found to reduce under and over-payment under capitation by up to two thirds, when 25% of health scheme expenses are reimbursed retrospectively (Keeler et al 1998).

(iii) **Service-Specific Risk Sharing**

Service-specific risk sharing strategies (carve outs) identify certain services and areas of health expenditure, and remove them from the capitated budget. Services to be carved out are generally
identified as those over which health schemes have the greatest incentive to cream skim, typically including pharmaceuticals, mental health services, dental services and neonatal services. The rationale for carving out these services is that their costs are more predictable, and therefore provide greater potential for cream skimming (Van de Ven and Ellis 2000). Carved out services may, or may not, be covered under supplementary insurance policies offered by the scheme. However, services may also be carved out that have highly unpredictable costs, representing a source of significant financial risk for the health scheme (eg heart transplants, intensive care).

7.4 Implementation Strategies for Risk Management

In the majority of countries the implementation of risk adjusted capitation schemes has been extremely cautious, particularly where capitation has replaced historic budgets. The purpose of this caution has been to enable schemes to manage the financial risks associated with the move to a new payment mechanism, particularly through minimising the financial dislocation to health schemes from changes to budgets. A range of implementation strategies may be used.

The dissemination of shadow (notional) budgets calculated under capitation to health schemes prior to its implementation informs schemes of future budgets to be expected under the new funding model. This allows health schemes to develop and implement strategies to manage risk and potential financial dislocation.

Capitation may guarantee that no future budget is lower in real terms than under the previous payment model. This may be achieved by the gradual phasing in of capitation budgets over a number of years, where health schemes that “lose” under the new system do not receive additional funding and health schemes that “gain” receive growth payments to gradually bring them up to the required capitated budget level.

Similarly, many capitation models use a movement to target approach. Under this approach capitation budgets are calculated for several years in advance and the distance from a target future budget allocation is determined for each health scheme. Strategies for moving each health scheme towards its target budget may then be developed, usually through bilateral negotiation, and ceilings may be placed on the rate at which schemes converge on their target.
8  Overview of Capitation Issues in Australia

In order to progress the debate about capitation funding in Australia - the prediction of expected costs of health care, risk adjustment methods etc., there are a number of issues that require attention. This paper has endeavoured to address the main issues in risk adjustment methods for predicting health service costs to contribute to a dialogue about health service funding and health system reform in Australia. This Section aims to bring together the discussion about risk adjustment methods and capitation funding to highlight key issues in its application in the Australian context. We suggest some preliminary conclusions can be drawn at this stage, but also some research questions around which the debate has only just started. Both are briefly described here. The section first presents conclusions in the general context of the Australian health system, followed by a brief discussion of risk adjustment methods and issues specific to the Australian Coordinated Care Trials. The section concludes with some areas for further debate and research.

8.1  Capitation Funding and Australian Health Services

Capitation funding, where health schemes receive funding commensurate with the expected health service costs of their enrollees, is relatively widely used across the world. It is also used in a number of contexts in Australian health services. The most notable example of capitation funding is the New South Wales resource distribution formula, which determines funding levels for Area Health Authorities. Other examples include the basis for some payments made to the States by the Commonwealth through the Health Care Agreements, and a range of individual health service programs funded at the Federal and State levels, notably the nursing home program.

Capitation funding may be a desirable option for devolving health service budgets for a number of reasons. Capitation provides an effective mechanism for capping budgets to secure cost containment. Where competition is allowed between health schemes, capitation can be used to reduce the potential for adverse selection and cream-skimming through the use of risk adjustment methods to predict health services costs. Irrespective of whether or not competition between health schemes is allowed, capitation can be used to pursue equity objectives, through the use of risk adjustment methods. Finally, if capitated budgets are used to break down traditional budget boundaries, capitation funding may be integral to moves to improve coordination and integration of care, with the potential to improve cost-effectiveness in service delivery and allocative efficiency in the health service mix. Thus, capitation funding has the capacity to provide significant benefits and meet health service objectives in a number of areas.

For example, at present casemix funding in many States, is used primarily to determine throughput volumes against hospital budgets, determined essentially on historic budgeting rules. As discussed in Section 2, historic budgeting tends to promote existing inefficiencies and inequalities between health schemes. Capitation funding could be employed, in this context, to calculate acute sector budgets based on the expected costs of health services reflective of the underlying health service needs of a population served by particular hospital(s).
Payment for individual services could still be made on the basis of DRG case payment, in the pursuit of efficiency in service delivery. In this way capitation funding and casemix funding could be combined to more effectively pursue health system objectives as they relate to cost containment, efficiency (both technical and allocative), and equity.

There exist a wide range of other contexts in which capitation funding may be beneficial, the majority of which are not described within the context of this paper. However two further examples are worth some attention.

In the context of the Australian Coordinated Care Trials, the development of risk adjusted capitation formulae is a fundamental requirement for Funds Pooling, a central feature of the Trial model. Predicting the expected costs of health services for enrollees is especially important because the financial viability of Trials is dependent on Trial managers being able to cover the costs of care from the Funds Pool. Further, if the Trial becomes insolvent, a range of other objectives will not be met. The use of historically determined budgets, based purely on the average costs of treatment in previous years, may expose Trials to a significant risk of becoming insolvent. This is because simple historic budgeting approaches do not properly account for differences in personal characteristics and organisational characteristics that influence enrollees’ expected health service costs. This requires a system of risk adjusted capitation. Otherwise enrollees may be refused access to services through the trial based on their characteristics, quality of care may be reduced, or the Trial may become insolvent.

Under moves to an integrated care model, many existing budget boundaries may be removed, (as in the UK Primary Care Group model). The broad aim of this type of health system model is to remove artificial budget boundaries to promote the cost-effective use of resources across a wider range of health services. This model attempts to promote allocative efficiency within a more holistic population health focus. To do so capitation funding may again be pivotal. By allocating funds to health schemes on the basis of enrollees' need for health service resources, risk adjusted capitation funding would form a more rational basis for the pursuit of allocative efficiency.

In the wider context of health service reform, risk adjusted capitation funding is a fundamental plank in the development of managed competition and/or integrated health care models. Managed competition model implies competition for enrollees by health schemes. This require a capitated payment per enrollee, be allocated to schemes. As outlined in Section 6, this capitated amount would have to be adequately risk adjusted to reduce incentives for adverse selection and cream skimming. Risk adjustment methods may be used to address equity objectives under managed competition, or under non-competitive health scheme models.

8.2 Risk Adjustment Methods and Australian Health Services

The main focus of this Paper has been on describing and discussing risk adjustment methods for predicting health service costs. Much of this description and discussion has focussed on the validity of alternative approaches to risk adjustment.
Some general conclusions and research issues on the validity of alternative approaches in the Australian context are presented below. However, this discussion really requires prior consideration of health system objectives.

(i) **Objectives and Supply Side Variations in Cost**

A robust debate about the objectives of the health system and the possible contribution of risk adjusted capitation to those objectives is desirable. It is of particular importance to ascertain the views of policy makers and the community about the relative importance of efficiency, equity, and access. While sometimes equity and efficiency objectives will be complementary, there may also be conflict between goals requiring choices and trade-offs. Such trade-offs require value judgements, which cannot be determined through regression based risk adjustment modelling. This matter is central to the development of an appropriate capitation rate. It is necessary to determine what it is designed to do. Is it merely to replicate past expenditures and historic access to services or is it to be used to redistribute resources to ensure a fairer access to services, however that is defined?

For instance this will be critical in deciding how to treat supply side influences on health service use and cost. Risk adjustment methods have adopted different stances in their treatment of avoidable cost differences between health schemes and service providers. These are supply side factors endogenous to the health scheme – within the scheme’s control – including managerial efficiency and local policies and practices. Essentially risk adjustment may seek to include or exclude these sources of variations in costs from capitation funding. Those health systems that have included avoidable cost differences in capitation payments have done so primarily to reduce the potential for risk selection by health schemes. Those health systems that have excluded avoidable cost differences have done so primarily in the pursuit of equity objectives.

Each approach may be valid for the stated objectives of capitation funding in alternative contexts. A discussion is therefore required in the Australian context about the objectives of the health system and any potential developments in capitation funding. A number of options exist that may represent valid objectives and approaches to avoidable cost differences. The primary objective of risk adjusted capitation may be to pool funds, in the pursuit of gains in allocative efficiency from service substitution. Equally, risk adjusted capitation may seek to reduce any potential for risk selection and quality scrimping. In both cases it may be appropriate to include avoidable cost differences in capitation payments. However, capitation funding may also be used to promote equity in access to health services. In this case it may be appropriate to exclude avoidable cost differences.

Experiences from overseas suggests that these, and other, potential objectives require considerable debate and research before a robust capitation model can be implemented. Research issues include the extent to which these objectives are relevant to the Australian context, the degree to which objectives are complementary, what trade-offs between objectives may be required, and the development of risk adjustment methods appropriate for any given set of objectives.
(ii) **Risk Adjustment Methods**

Simple historic expenditure approaches to budget setting - where future budgets are based on current utilisation and incremental adjustments that do not account for differences in enrollees health care needs - will perpetuate and promote inefficiency and inequities in the health system. Even if the primary goal of budget setting is the creation of the budget itself, it is unlikely that the incentives produced by historic expenditure approaches will be acceptable to policy makers and the public. Sound risk adjustment methods are required.

Overseas risk adjustment models cannot be directly imported because health expenditures in Australia reflect a different set of health system delivery and finance methods than found elsewhere, and the relationship between needs, supply and expenditure in Australia will differ from that found overseas. Detailed analysis is required in the Australian context, using Australian data.

Poorly conceived risk adjustment methods have typically failed overseas. Implementation in such circumstances has rarely been achieved. The major advances in risk adjustment have been in the US, the UK and the Netherlands. In these countries risk adjusted capitation models have been successfully implemented, with efficiency and equity gains. This has only been possible through significant investment in research to develop sound models and methods. Where this investment has not been made capitation models have enjoyed relatively limited success.

(iii) **Individual Level versus Aggregate Level Analysis**

Risk adjustment models have been based either on individual/patient level data or small area data (such as census collector district or postcode). Typically small area data has been used in the non-competitive health scheme models. This is primarily because individual level data is less commonly found in health systems where individual insurance claims data is not required. However, under a non-competitive model, it is also less critical to establish a correct capitation weight for each individual. There is limited potential for risk selection where enrollment is dictated by factors outside health scheme control (such as geographic location), and budgets are typically set for a defined population group.

In theory, the use of individual level data is preferable, regardless of health scheme type. Individual level analysis allows for more precise estimation of individual’s expected costs, because it allows more accurate description of the relationship between health service use and costs, and individuals’ demographic, health and social characteristics. Small area data does not capture the full extent of heterogeneity between individual’s within small areas, and may result in less precise predictions of costs at the individual level. Aggregate level summary measures of need in a given population do not necessarily reflect the sum of the needs of the individuals that go to make up that population. This is often referred to as the ecological fallacy - inferences made from aggregate level analysis may not hold at other levels of aggregation. More specifically, in the context of risk adjustment, relationships found between needs factors and utilisation/costs at the aggregate level may not hold at the individual level. This may cause problems in the use of aggregate level models for the risk adjustment of capitation payments.
Small areas may not consist of homogenous individuals. If there is considerable variation within small areas, in terms of population characteristics, as well as variation between small areas, inferences from small analysis to lower levels of aggregation may be inappropriate. Essentially, aggregation of the data may result in some relationships between individual level needs factors and expenditure being masked. More importantly, the purpose of risk adjustment is to estimate the link between individual’s needs and health expenditure. This holds even if payments are made at the health scheme level (rather than at the individual level) because resources allocated at the scheme level are still intended to lead to the most appropriate patient care for individuals.

The gold standard approach to risk adjustment should therefore be to utilise individual level data wherever possible. Furthermore, the most appropriate design would be to analyse a large cohort of individuals to robustly estimate relationships between needs factors and expenditure over time.

(iv) Capacity of Models to Predict Health Service Costs

The capacity of models to predict health service use and cost depends primarily on the structure of the model, the choice of explanatory variables, whether individual or small area data is used, and the scope of services covered. Whether the capitation formula is consistent with the financial survival of health schemes will depend on the extent to which unexplained variation is a reflection of random error (rather than statistical bias), scheme size in terms of number of enrollees, and strategies for risk sharing.

Using individual level data, it has been estimated that capitation models can predict up to 20-25% of variation in health service costs. This is widely held to represent the proportion of health service costs that is potentially explicable under statistical analysis. The remaining 75-80% of variation in costs is due to random factors. In practice, many capitation formulae based on individual level data have reported that 10-12% of the variation in costs was explained in regression modelling. Studies of selected types of health services have demonstrated less random variability in costs for some services, for example chronic care and pharmaceuticals, and have reported as much as 56% of variation in costs explained in modelling.

In relation to financial survival of health schemes, even risk adjustment models which explain only 10% of individual level variation will be likely to cover actual costs, provided the enrolled population is of sufficient size (and budgets reflect predicted cost). Whilst little empirical work has been carried out on optimal scheme size for capitation formula based on individual level data, an enrolled population of 5,000 is suggested as sufficient to minimise the risk of financial insolvency due to random influences on costs.

The percent of variation in health service cost predicted by models using small area data tends to be higher, at around 55%. This is not because of any superiority in the model structure, but because the combination of data, on the cost side as well as in relation to independent variables, removes substantial variation in the data. However, in translating to financial robustness of schemes, far larger enrolled populations are required.

At an enrolled population of 10,000 persons it has been estimated that approximately 17% of Schemes would find expenses more than 10% in excess of capitation payments, (where small areas used in analysis have an average population of 10,000). This proportion falls rapidly to 0.1% for an enrolled population of 100,000 persons.
(v) **Modelling and Explanatory Variables**

Variables used in risk adjustment models are drawn from a range of measures for needs factors and supply side factors.

Typical needs factor variables include:

- **Demography**
  - Age, Sex, Ethnicity

- **Employment/Disability Status**
  - Employed, Unemployed, Pensioner,
  - Temporarily unable to work, Permanently sick

- **Health Status**
  - Self reported/survey-based morbidity, Mortality rates
  - Permanent disability/dependency status, Low birth weight
  - Previous inpatient/outpatient diagnosis (DCG diagnostic cost group,
    ACG ambulatory care group etc)

- **Socioeconomic Status**
  - Homelessness, Marital status, Income, Housing tenure
  - Socioeconomic Status/Social class, Education level
  - Elderly living alone, No carer in household

- **Geographic location**
  - Region/area of residence, Rurality, Urbanisation
  - Remoteness, Population Density

Typical supply side factor variables include:

- **Provider Characteristics (Exogenous to health scheme)**
  - Input prices, Labour prices, Capital rental prices
  - Location specific service delivery costs
  - Rurality, Urbanisation, Remoteness

- **Provider Characteristics (Endogenous to health scheme)**
  - Market power - ability to get price discounts
  - Insurance coverage - deductibles, co-payments etc
  - Contracting arrangements, Practice style
  - Health scheme management policies and practices

Prior diagnosis information may represent a particularly important needs factor for exploration in future research. This could be achieved by initially importing the Boston DCG framework, and developing it in the Australian context (in much the same way DRGs were imported). It is highly unlikely that DCGs will be directly transferable because of the significant differences between the Australian and US health systems.

A wide range of potential needs factors should be considered in any analysis. Competitive health scheme approaches to capitation have been weakened by their failure to consider broader determinants of health and health needs.
Non-competitive approaches have been much more thorough in their treatment of a range of needs factors. Modelling techniques should include consideration of both the potential interrelationships between supply, demand and expenditure, and the hierarchical nature of health service organisation and delivery. This requires sophisticated approaches to regression modelling, including the use of multilevel modelling techniques.

(vi) **Population and Service Coverage**

Little attention has been given to the scope of services to be included in health schemes. However, this has significant implications for the development of risk adjusted formulae and incentives for cost shifting. In general, the narrower the scope of services covered, the greater the capacity for cost-shifting to services not included in the formula. But where quite distinct service types are to be covered, (such as acute care, mental health, primary care, residential care), it is probable that a number of distinct capitation formula will need to be developed.

In particular, the choice of a regional population based model or a targeted high risk enrolled population, has fundamental implications for the approach to capitation, and the likely robustness of the model. There are arguments both ways. The potential advantages of a population-based approach include:

- greater certainty in numbers to whom the capitation rate is to apply;
- the possibility of developing a generic capitation based funding formula that can be applied across the nation, rather than relating only to a unique enrolled population;
- greater opportunity to introduce system wide change;
- greater opportunity to introduce prevention approaches to care;
- reduced opportunities for cost shifting outside the scheme;
- possibility of funding at the regional level rather than via individual agencies;
- there is no opting out, and no issues of either cream skimming or individuals (eg those now incurring high costs and not wanting to be monitored) selecting not to participate;
- the model would not offer choice of scheme, but could potentially encourage competition amongst providers;
- facilitate a strong health services planning focus to ascertain community preferences, relative access to services, capacity of the system to deliver best practice/ cost effective care.

On the other hand there are a number of difficulties with moving to a population-based model and some advantages of a targeted approach. These include:

- reduced establishment costs, the possibility of implementing a larger number of smaller schemes, all of which can provide important lessons;
- the possibility of gaining informed consent to gather patient level data, which might not be possible in a population based approach, just by virtue of the numbers involved;
- the possibility of targeting particular high needs groups that may most benefit from care coordination/integrated care;
- with a defined target it is possible to enroll a control population which provides a means for checking the viability of the budget calculation and assessing the impact of the health funding and delivery model on health services use and cost and health outcomes.
(vii) Approaches to Risk Sharing

Strategies for risk sharing using blended payments (a mix of capitation and retrospective reimbursement) are an effective tool for managing financial risk to budget holders. Outlier and proportional risk models provide some protection against insolvency for budget holders. Condition-specific and service-specific risk sharing may also be useful strategies, but evidence is sparse for these arrangements.

What also seems clear is that provided the population is of a reasonable size, which might be as low as 5,000 persons (or even less), random variation may not be excessive. Reduction in risk through, say the exclusion of certain service types from the pool and the capitation rate, may be unnecessary, and contradictory to the achieving resource shifts between programs. The financial viability health schemes is likely to be influenced more strongly by an incorrect capitation rate, high administration and management costs, and cost pressures outside their control.

In an Australian context the existence of services over which the scheme manager has no control, notably access to MBS, PBS and to a lesser extent public hospital services, represents an important issue that needs to be addressed. Similarly how to manage costs pressures, such as the relative use of private hospitals and thus private fee-for-service doctors which are outside the control of the scheme manager needs to be debated.

8.3 Risk Adjustment and the Australian Coordinated Care Trials

(i) Aims of the Australian Coordinated Care Trials

The Australian Coordinated Care Trials arose primarily to address the management of persons with complex chronic conditions. There was a concern that traditional health funding and delivery arrangements were unable to deliver the optimal mix of services, with the necessary coordination of care, given the myriad agencies, service providers and separate programs used. The creation of a single funds pool was seen as a means to break down individual program boundaries and facilitate resource shifts across programs and agencies. The introduction of care coordination activities was a means to achieve greater continuity in care and improve the quality of care. A primary concern of the Trials was the extent of cost shifting which was seen as wasteful in terms of administrative effort, and potentially distorting of the health service mix.

The primary driver of the coordinated care trials was thus efficiency. However, the Coordinated Care Trials also provide an opportunity to document differences in access to services, to determine reasons for these differences, and to consider the implications of moving to a system in which access is largely determined by need. Numerous separately funded programs also create inequities between those eligible for and able to access particular programs and those who miss out. Funds pooling, in bringing together a larger number and wide range of consumers into a ‘single program’, provides a potential to address equity and access objectives.

* For example, in the population based trials in with ATSI communities, the HIC contribution to the Funds Pools was substantially greater than ‘expected cost’, in recognition of the historic under-funding of private medical and pharmaceutical services to these communities.
A final consideration in establishing a Funds Pool is the requirement for budget neutrality and financial viability. The Funds Pool should in theory, provide a means to monitor the impact on total resource use of changes, and guarantee that the health system reform would proceed in the context of ‘current resources’. However, the Funds Pool will only be an effective means for ensuring budget neutrality if financial flows to and from the pool are a true reflection of real resource flows. This requires a capacity to estimate ‘expected health service use’, to set an appropriate capitation rate and to obtain contributions from ‘all’ providers and agencies at that rate.

(ii) Nature of the Funds Pooling Model for the CCTs

The funds pool model adopted by the coordinated care trials is quite different to the models common overseas (US, UK and Europe), to which most of the capitation literature is addressed. The difference is fundamental and lies in the source of funds for the pool. Under most capitation models, funds are contributed by a single funder (or at most two or three), usually the national government, or private health insurer. The Scheme manager typically negotiates with providers about the purchase of services, but not with providers (or even funders) over the contribution of income or revenue to the Fund. The contribution rate is developed by the funder not the Scheme manager, and is designed to reflect ‘individual risk/need’ of those that are, or could be, enrolled. The contribution rate is designed to avoid consistent bias to minimise incentives for cream skimming (where there is choice about buy-in) and to achieve equity between Schemes and populations.

However, under the Australian coordinated care trials, the Scheme manager must obtain funds from the programs and agencies, from revenues they have received from the funders. There are thus two separate requirements in setting up a robust Funds Pool, firstly to ascertain a technically correct funding/capitation rate, and secondly to negotiate with the numerous agencies to contribute at the technically correct rate.

This creates two major hurdles that are unique. In calculating the capitation rate, rather than a single (or perhaps three or four) risk adjusted formulae being developed by the funder, current program boundaries dictate the need to develop numerous separate risk adjusted models. This contrasts with a regional based funding model in which the primary funders, the State and Commonwealth Health Departments would estimate the total costs of each major program (say acute care, residential care, HACC) for the region, requiring say four formula, rather than dozens.

Secondly, even if a robust capitation formula can be developed for each agency, there is no guarantee the agency would be persuaded to contribute at that rate. In fact, as the explicit aim is to achieve resource shifts, to purchase care co-ordination services from the Pool and meet additional administrative and management costs from the Pool, agencies may have little incentive to contribute at the calculated rate based on the expected cost of usual care.

For example, separate risk adjusted rates will need to be developed for each enrolled population for private medical services funded through Medicare and other funders (eg Veterans Affairs), separate formula for each of dozens of community based agencies, for each community health centre, for each public hospital (or network), each private health insurance fund, for each residential care facility.
There are thus two separate issues: to ascertain a technically correct funding rate and to negotiate contributions to the pool at the technically correct rate from numerous agencies. A logical alternative is for the primary funders to directly contribute to the Scheme, withdrawing an equivalent amount from each agency to cover the additional administrative and management costs of care coordination.

The distinction is illustrated in Figure 8.1 below, where the Australian model is seen to be far more complex than those typically found overseas.

**Figure 8.1** Funds Pool model, typical models and the Australian CCTs

---

**Typical Overseas Model**

- **Funder**: Govt and/or private insurer
- **Health scheme**: Health Authority/HMO etc
- **Providers**: \( p_1, p_2, p_3, \ldots, p_n \)
- **Funds**: $$$

**Australian Coordinated Care Trials**

- **Funder**: Govt and/or private insurer
- **Health scheme**: Trial Management
- **Providers**: \( p_1, p_2, p_3, \ldots, p_n \)
- **Funds**: $$$

---

### 8.4 Research Questions for Australia

There is a substantial world-wide research agenda to develop risk adjustment methods to predict the expected costs of health services. These methods are used to develop risk adjusted capitation funding models for health services.
A rigorous approach to the research task is critical, in order to engender confidence in the resulting formula in terms of its robustness, transparency and ability to provide correct incentives. Inappropriate risk adjustment methods can result in insolvency for health schemes, or in the competitive context, perverse incentives for risk selection (cream skimming and adverse selection). It is hoped that this Paper will contribute to an on-going dialogue about health system funding and delivery and the role for risk adjustment and capitation funding in Australia. The types of issues that will need to be considered include:

(i) **Health System Objectives and the Role of Capitation Funding**

As noted above, a robust debate about the objectives of the health system and the possible contribution of risk adjusted capitation to those objectives is desirable. It is necessary to determine the role of capitation and in particular whether equity as well as efficiency is to be addressed. Research issues include the relative importance of the possible objectives, how they are conceived by the Australian community, the degree to which objectives are complementary, the trade-offs between objectives that may be acceptable and the development of risk adjustment methods appropriate for a given set of objectives.

(ii) **Supply Side Factors and Risk Adjustment**

Risk adjustment methods have differed in their treatment of avoidable cost differences - supply side factors within the scheme’s control, including managerial efficiency and local policies and practices. Risk adjustment may seek to include or exclude these sources of variations in costs, from capitation funding. Those health systems that have included avoidable cost differences in capitation payments have done so primarily to reduce the potential for risk selection by health schemes. Those health systems that have excluded avoidable cost differences have done so primarily in the pursuit of equity objectives.

A discussion is required in the Australian context about the objectives of the health system and the implication for the treatment of variation in cost attributable to supply, in relation to variables within and outside the control of the Scheme. A number of options exist that may represent valid objectives and approaches to avoidable cost differences. The primary objective of risk adjusted capitation may be to pool funds in the pursuit of efficiency gains from service substitution. Equally, risk adjusted capitation may seek to reduce any potential for risk selection and quality scrimping. In both cases it may be appropriate to include avoidable cost differences in capitation payments. However, capitation funding may also be used to promote equity in access to health services. In this case it may be appropriate to exclude avoidable cost differences.
(iii) **Population and Service Coverage**

Decisions need to be made about the services to be provided by health schemes and therefore included in risk adjusted capitation formula. Options include for instance, primary medical care only, all primary care, primary plus secondary, plus selected or all tertiary care, plus residential care and may include selected components of disability and welfare services. These decisions should reflect the potential for improvements in allocative efficiency through service substitution from the removal of budget boundaries.

(iv) **Capped or Open-Ended Budget/Treatment of Private Health Insurance**

In adopting a risk adjusted capitation based model of funding there is a presumption that health service costs can be constrained by the scheme management. Costs which are outside the control of scheme management, might be determined outside the capitation arrangement. This might include regional based supply side variables such as input prices, or travel costs associated with distance. They might also need to accommodate national commitment to unrestricted access to certain types of services, such as to private medical services included on the MBS which cannot be the subject of restriction. Similarly conditions of access to public hospitals are dictated by joint Commonwealth State Agreements. Treatment of services funded through private health insurance is also not necessarily within the control of potential scheme managers. How these issues can be, and are to be, reconciled within a risk adjusted capitation funding model need to be explored.

(v) **Support for Research Program to Develop Capitation Formula**

The wider adoption of capitation based funding is an important possible future direction for the Australian health system. Progress in this area relies heavily on two types of research:

- into the exploration and development of risk adjusted capitation funding models
- into the characteristics of the Australian health system, and how or whether they can be reconciled with risk adjusted capitation funding

The three countries which have led the way in developing risk adjustment methods, and in successfully implementing capitation funding models – the US, the Netherlands and the UK – have all devoted significant time and resources to these key research tasks. Experiences have been remarkably similar in all three countries: that a core group of 3-5 researchers working over a period of 2-3 years is required as a minimum, to develop methods and analysis to construct robust risk adjusted capitation formulae that provide appropriate incentives consistent with health system objectives.
References


Carr-Hill R 1987, Health Status resource allocation and socio-economic condition, University of York, Centre for Health Economics, York.


Curtis SE 1990 ‘Use of survey data and small area statistics to assess the link between individual morbidity and neighbourhood deprivation’, *Journal of Epidemiology and Community Health*, 44:62–68.


Lamers LM 1999b ‘The simultaneous predictive accuracy for future health care expenditures of DCGs, PCGs and prior costs’, paper presented at the iHEA-Conference, 6–9 June 1999, Rotterdam.


Van de Ven WPMM & van Vliet RCJA, Schut FT & Van Barneveld EM 1997, *Access to coverage in a competitive individual health insurance market: via premium regulation or via risk adjusted vouchers?* Department of Health Policy and Management, Erasmus University, Rotterdam.


Van Vliet RCJA & Lamers LM 1998, ‘The high costs of death: should health schemes get higher payments when members die?’, *Medical Care*.


