Capitation payments to competing Dutch sickness funds based on diagnostic information from prior hospitalizations

Capitation payments to competing Dutch sickness funds based on diagnostic information from prior hospitalizations

Normuitkeringen voor ziekenfondsen gebaseerd op diagnose-informatie van ziekenhuisopnamen in het verleden

Proefschrift

Ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam op gezag van de
rector magnificus prof. dr P.W.C. Akkermans M.A.
en volgens het besluit van het college van promoties.
De openbare verdediging zal plaatsvinden op
woensdag 3 december 1997 om 13.45 uur

door

Leida Maria Lamers
geboren te Rotterdam

Promotiecommissie

Promotor:	prof. dr W.P.M.M. van de Ven
Overige leden:	prof. dr J.P. Mackenbach
	prof. dr D. Post
	prof. dr E. Schokkaert
	dr R.C.J.A. van Vliet (tevens co-promotor)

ISBN 90-9010833-5

Cover: Tanja E. Algra, Bureau Karakter, Delft
Printed by: Ridderprint, Ridderkerk

To my parents

Contents

1

Introduction

1.1 Background

In many countries market-oriented health care reforms are high on the political agenda. The purpose of these reforms is to make resource allocation in health care more efficient, more innovative and more responsive to the consumers' preferences. The Netherlands is no exception in this respect. The Dutch health care reform shows close similarities with the reforms in, for instance, Belgium (Kesenne, 1996), Germany (Graf von der Schulenburg, 1994; Files and Murray, 1995), Switzerland (Beck and Zweifel, 1996), Israel (Chinitz, 1994) and the U.S. (Newhouse, 1994). A common element of these reforms is that the consumers may choose among competing health insurers or health plans, which are largely financed through premium-replacing capitation payments.

In a competitive health insurance market premiums will reflect the risks of the insured persons. Risk rated premiums are socially undesired, because they may hinder access to (good quality) health care for high risk individuals (Light, 1992). A remedy to this problem are regulations that would prohibit insurers from denying coverage or charging different premiums to different individuals based on their health status. Under these regulations an insurer must charge each potential beneficiary the same premium for a given insurance plan. With such a system of 'community rating', the healthy subsidize the cost of care for the less healthy. For insurers, community rating will create financial incentives to attract only healthier clients because these people will (on average) pay more in premiums than they will generate in claims. In such a context insurers' profits will depend more on the plan's ability to contract healthier beneficiaries and less on efforts to provide high-quality service at the lowest

price (GAO, 1994a). A key strategy for insurance market reform is to provide incentives for health plans to compete for enrollees by offering better quality and efficiency rather than practising risk selection. Failure to address risk selection will continue to have serious consequences both for access to care for vulnerable populations and for the financial viability of health plans (Gauthier et al., 1995).

Under a reform model based on community rating and competing health plans, risk-adjustment methods will be necessary to compensate for differences in relative risk across capitated plans or provider networks. Without adequate risk-adjusters, entities bearing risk have a financial incentive to select or market to healthier enrollees. Under perfect risk-adjustment, individuals pay the same premium regardless of their health status, but the premium payments received by the insurers are adjusted to reflect the differences in individuals' expected costs (PPRC, 1994a; GAO, 1994a).

The major goal of a risk-adjustment mechanism is to set the proper incentives for both consumers and competing health plans. A risk-adjustment mechanism removes the effect of how sick other members are from the health plan price. When consumers face health plan prices free of risk, they choose a health plan because of its efficiency, services and (perceived) quality. Plans with higher risk members should receive more money from the risk-adjustment mechanism than plans with lower risk members. In the ideal case, plans are indifferent to the health risk of their members or prospective members. With a sufficiently good risk-adjustment mechanism, some health plans should be willing to specialize in treating specific high-cost diseases (Bowen, 1995). However, the development of such a risk-adjustment system appears to be a major technical problem. The purpose of this study is to contribute to the development of an adequate risk-adjustment mechanism. This study is conducted in the Netherlands using Dutch data. The health care reform in the Netherlands is described briefly in the next section [1] for a better understanding of the context for this study. In section 1.3 the system of risk-adjusted capitation

[1] Schut (1995) gives an extensive analysis of the Dutch health care reform.

payments in the Netherlands is described and is followed by a section explaining the purpose of the study.

1.2 Health care reform in the Netherlands

In 1988 the Dutch government proposed fundamental changes in the structure and financing of health care. The principal aim of the proposals was the creation of a health care system which combined an essentially social character with effective mechanisms that guarantee cost containment and efficiency. The main elements of the proposals were a restructuring of the insurance system and a greater role for market elements in the health care system (WVC, 1988).

Regulated competition among insurers as well as among providers was a crucial element of the reform. The Dutch model could be considered the first attempt at a nation-wide implementation of Enthoven's (1978) Consumer-Choice Health Plan. The proposed system could be best characterized as a compulsory national health insurance ('basic insurance') based on regulated competition.

In the Netherlands three types of health insurance for noncatastrophic risks (hospital care, general practitioners, inpatient and outpatient specialist care etc.) can be distinguished. Approximately 60 percent of the population in the lowest income brackets are compulsorily insured by a sickness fund. Non-governmental employees, pensioners and social security beneficiaries (and their dependents) with incomes below an annually determined income level are members of a sickness fund. Provincial and municipal civil servants, which account for about 6 percent of the population, have a separate mandatory health insurance arrangement. One-third of the population, including higher-income employee groups, the self-employed and state government officials have private health insurance. Although private health insurance is voluntary, in 1992 only 0.7 percent of the population was uninsured (CBS, 1993).

In addition to the basic insurance for noncatastrophic risks, optional supplementary insurance may cover those facilities not included in the basic insurance, e.g. dental care for people over 18 years and extended physiotherapy.

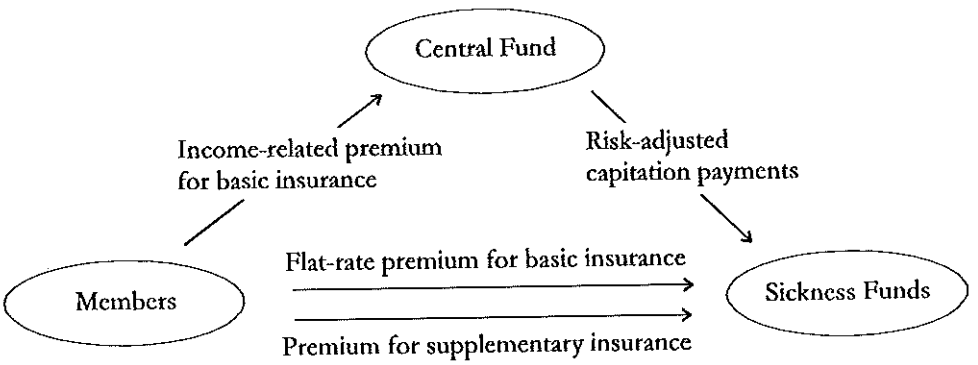
For catastrophic risks there is a mandatory national insurance program: the Exceptional Medical Expenses Act (AWBZ). This act covers social services and exceptional medical expenses, such as care for mentally and physically handicapped persons, institutional psychiatric care, nursing home care and other forms of long-term institutional care.

In contrast to former cabinets, the new coalition cabinet which came into office in 1994, decided that regulated competition among health insurers should be restricted to noncatastrophic risks (VWS, 1995). Catastrophic risks appear to be the least suited for a market-oriented approach, because there are no (or not enough) prudent buyers to motivate competing health insurers to provide efficient, high quality care (Van de Ven and Schut, 1994). The new cabinet aimed at a convergence of sickness funds and private health insurers by parallel reforms of both insurance sectors. This thesis will be restricted to the social health insurance sector, because in this sector risk-adjusted capitation payments were introduced in 1993.

All sickness fund members pay an income-dependent premium levied by the tax collector as well as a small flat-rate premium to be paid directly to the sickness fund of their choice (figure 1.1). Sickness funds are obliged to have an annual open enrolment period and to obey other procompetitive regulations. The income-related premiums are collected in a central fund, from which the sickness funds receive risk-adjusted capitation payments. A risk-adjusted capitation payment is independent of the chosen sickness fund and equals the predicted per capita costs within the risk group to which the member belongs, minus a fixed amount. The fixed amount is equal for all individuals and is about 10 percent of the average predicted per capita costs for the basic benefits package. The deficit thus created is met by the flat-rate premium paid by the member directly to the sickness fund of his or her choice. A sickness fund is obliged to quote the same flat-rate premium to all of its members who choose the same insurance option.

This premium will reflect the difference between actual costs and the risk-adjusted capitation payments thus creating an incentive for the competing sickness funds to be efficient.

Figure 1.1 Financing system for the Dutch social health insurance sector



The sickness funds are expected to function as an intermediary between consumers and providers of care. To a high degree, insurers and providers are free to negotiate conditions of contracts. Consumers are free to choose among different sickness funds. Premiums paid by members do *not* reflect their risks as they would in a completely free competitive market, thus guaranteeing equity. Premiums in principle reflect the efficiency and cost-generating behavior of the contracted health care providers. In this way it is expected that a situation arises in which:

- the consumers are being rewarded for choosing efficient sickness funds and choosing cost-effective providers of care;
- the providers are being rewarded for effective and efficient provision of care;
- the sickness funds are stimulated to be prudent buyers of care on behalf of their members by contracting efficient providers and by conducting market research to discover consumers' preferences.

During the last 50 years the sickness funds were fully reimbursed for all health care expenditures of their members. As part of the reforms described

above this retrospective reimbursement system was replaced in 1993 by a system of risk-adjusted capitation payments. In most regions there were only one or two sickness funds, so there was hardly any real consumer choice. Now all sickness funds operate nationwide. Since 1994 they have the option to selectively contract with providers of care. In 1997 relatively large differences occurred between sickness funds in the flat-rate premiums paid by the members directly to the sickness fund. So, within a short period of time sickness funds have been transformed from pure administrative bodies into risk-bearing enterprises.

1.3 Risk-adjusted capitation payments

Adequate risk-adjustment is critical to the success of the Dutch health care reform, as well as to the market-oriented health care reforms in other countries (Van de Ven and Van Vliet, 1992). The development of a system of risk-adjusted capitation payments (RACPs) is a major technical problem, which has been seriously underestimated in the Netherlands.

The capitation payments should be adjusted for the health care needs of the members. The payment per member is dependent on the risk category to which the person belongs. RACPs should account for predictable variations in annual per-person health care expenditures, as far as these are related to health status. The RACPs are intended to provide the competing sickness funds with an incentive for efficiency. However, when the RACP risk-groups are rather heterogeneous, the capitation system has two disadvantages. First, RACP risk-groups that are too heterogeneous may result in an unfair distribution of payments to the sickness funds. A sickness fund with relatively unhealthy members per risk group will be underpaid. For a sickness fund with relatively healthy members per risk group the opposite holds. A second disadvantage resulting from too heterogeneous RACP risk-groups is that cream skimming may be very advantageous to the sickness funds. Cream skimming (or preferred risk selection) is selection by the sickness fund of so-called preferred risks, i.e.

those persons for whom the sickness fund considers the expected costs to be (far) below their capitation payment, given the regulatory regime regarding the flat-rate premium.

As pointed out by Pauly (1984), cream skimming is the result of premium regulation. In a free competitive market sickness fund will differentiate their premiums according to risk. Premium differentiation may hinder access to health care for high risk individuals. Risk-adjusted capitation payments can be seen as a form of regulation that attempts to simulate the premium structure in a competitive insurance market without having the adverse effect of (extreme) premium regulation. However, cream skimming may occur if the sickness funds can distinguish several subgroups of individuals with different expected costs within a RACP risk group and if premium regulation prohibits them to relate their premiums to the relevant risk factor.

The adverse effects of cream skimming are threefold. Firstly, for the (chronically) ill the access to good health care may be hindered. Sickness funds have no incentive to invest in good (quality of) care for these people. On the contrary, sickness funds that do make improvements in this area will attract (chronically) ill resulting in financial losses if the risk-adjusted capitation payments are inadequate. Therefore, sickness funds will try to attract the preferred risks and deter the non-preferred risks [2]. Secondly, in the case of an insufficiently sophisticated RACP efficient sickness funds might be driven out of the market by an inefficient sickness fund that is successful in cream skimming. Thirdly, whilst individual sickness funds can gain by cream skimming, it only shifts costs to others, so there is no social gain. In fact, because of the costs involved in the process of cream skimming, there are only social welfare losses. In sum, if cream skimming takes place, it is counterproductive with respect to three supposedly positive effects of competition, i.e., improving the quality and efficiency of care and becoming more responsive to the consumers' preferences.

---

[2] Cream skimming can take place both at the enrolment of new members and at disenrolment. An overview of possible forms of cream skimming is given by Luft and Miller (1988) and Van de Ven and Van Vliet (1992).

In 1993 and 1994 sickness funds received RACPs that were based solely on age and gender. In 1995 the capitation formula was extended with region and disability. Disability indicates whether or not a sickness fund member is a recipient of disability benefits. However, these risk-adjusters are much too crude. To reduce the above-mentioned disadvantages of too heterogeneous risk groups, the government introduced a partial capitation system. From 1993 to 1995 the sickness funds were responsible for about 3 percent of the difference between their actual expenses and their predicted expenses based on age, gender, region and disability. The remaining 97 percent was retrospectively reimbursed. In essence this is a blended payment system as proposed by Newhouse (1994), where the weight on current expenditures is 0.97 and the weight on predicted expenditures is 0.03. Although this partial capitation system strongly reduces the problems of both (potential) cream skimming and an unfair distribution of payments over the sickness funds, it also strongly mitigates their incentives for efficiency.

In 1996 the hospital costs incurred by the sickness funds has been split into production-dependent and production-independent costs. The sickness funds are held financially responsible for the production-dependent hospital costs only, because they can influence these costs by negotiating with hospitals about the hospital production for their members. The production-independent hospital costs are largely retrospectively reimbursed (ZFR, 1995a, 1995b). In 1996 the financial responsibility of the sickness funds increased from 3% to about 15% of the difference between their actual expenses and the RACPs they received. In 1997 a system of excess of loss has been introduced, implying that for those members whose costs exceed a certain threshold 90% of the excess costs are retrospectively reimbursed. In 1997 the financial responsibility of sickness funds further has been increased to about 25%. The government intends to further reduce the partial capitation gradually (except for the production-independent hospital costs), resulting in abolition in 1998 (VWS, 1995). In order to provide sickness funds with complete financial responsibility, they will need more tools for improving efficiency and the risk-adjusted capitation payment system needs to be improved. This thesis focuses on improving the risk-adjusted capitation payment system by adding

diagnostic information from prior hospitalizations as a risk-adjuster to the capitation payment formula.

1.4 The purpose of the study and guidance for the reader

Demographic variables like age, sex, disability and region are too crude as risk-adjusters for capitation payments. The capitation system can be improved by extending the set of risk-adjusters with factors that are more directly related to health. In chapter 2 an overview of the literature about potential risk-adjusters is given. One of the most promising risk-adjusters seems to be prior utilization in combination with diagnostic information. This thesis aims to contribute to the improvement the risk-adjusted capitation payment system by extending the set of demographic risk-adjusters with diagnostic information from prior hospitalizations.

Chapter three describes the data sources used for the empirical analyses presented in chapters four to nine, as well as the method used throughout this thesis to evaluate the predictive accuracy of capitation models.

Diagnostic information from prior hospitalizations can be incorporated in a capitation model in the form of Diagnostic Costs Groups (DCGs). The DCG model was developed with, and applied to, data from the US Medicare program, which mainly covers people of 65 years and over. Whether this DCG model can be applied to Dutch sickness fund data of persons of all ages [3] is studied in chapter four. For the development of the DCGs, diagnoses from prior hospitalizations were first classified into 78 clinically homogeneous groups, which were further clustered into 9 groups according to the empirically determined similarities in the *future* costs of individuals hospitalized with different diagnoses. In chapter five the

[3] A version of chapter four in combination with chapter eight has been published in Medical Care (Lamers and Van Vliet, 1996).

development of a DCG classification based on Dutch follow-up costs is described.

Chapter six deals with the question of how to treat so-called 'high discretion' diagnoses in a DCG model. A DCG capitation model should not reward hospital admissions for diagnoses for which the decision to hospitalize may involve high levels of discretion. Persons with hospital admissions for high discretion diagnoses should be treated as persons without a hospitalization. In chapter six high discretion is defined in the Dutch context and the effect of removing high discretion diagnoses from the DCG classification on the predictive accuracy of the DCG model is studied.

One of the requirements a risk-adjuster should ideally meet is validity. There are two conditions DCGs should meet to be a valid risk-adjuster. First, DCGs should have the ability to predict future health care costs. This refers to predictive validity or criterion validity, which is studied in the chapters four to six, where the ability of several variants of the DCG model to predict future health care expenditures is examined. The second condition to be met is that DCGs measure (semi-)permanent health. This second condition refers to construct validity. In chapter seven the construct validity of DCGs is examined.

Only about 6.5% of the sickness fund members has one or more hospital admissions in a year. Every year there will be persons in poor health with predictably high expenditures, who are not hospitalized and thus not classified into a DCG. Therefore, the usefulness of incorporating DCGs over a longer base-line period in the capitation model is studied in chapter eight. This leads to a multi-year DCG model.

In chapter nine the predictive accuracy of the multi-year DCG model is evaluated using health survey data [4]. This chapter deals with the question of which subgroups of sickness fund members the DCG model is able to predict health care expenditures accurately and which subgroups are missed

[4] A version of chapter nine is submitted for publication.

by the DCG model. To what extent incorporating information about these missed subgroups in the capitation model is complementary to the diagnostic information from prior hospitalization is examined.

Finally, the main findings are summarized and discussed in chapter ten.

2

Risk-adjusters for capitation payments

2.1 Introduction

The process for risk-adjusting payments to health plans involves three steps:

- Identifying the risk-adjustment variables. As risk-adjustment variables or risk-adjusters, a variety of observable factors can be used to predict the future costs of health care for each individual;
- Estimating the relationship between each of these factors and the cost of appropriate health care;
- Using these estimated relationships as a basis for determining the payments per enrollee and transferring funds to insurers (GAO, 1994a).

The first two steps in this process are also referred to as risk assessment (Bowen, 1995).

The crux of any comprehensive risk-adjustment system is the method used to measure the level of 'risk' -the underlying health care needs- of the beneficiary group in question (PPRC, 1994b). Over the last decade, methods for measuring relative risk across beneficiary groups received considerable attention, much of which has been directed at the AAPCC (adjusted average per capita cost) used by the Health Care Financing Administration (HCFA) to pay its risk contract Health Maintenance Organisations (HMOs) in the United States. These risk contracts with full prospective payment for enrolment of Medicare beneficiaries intend to be priced on the basis of 95 percent of what HMO enrollees would have cost Medicare had they remained in the fee-for-service (FFS) sector. This requires a methodology for estimating the hypothetical cost to Medicare. The current HMO payment model, known as AAPCC, starts with

projected Medicare FFS reimbursements per capita in counties of residence of HMO enrollees as the basis for payment. The average cost projections are then adjusted for differences in the distributions of HMO enrollees by age, sex, welfare status, and institutional status relative to the distribution of beneficiaries in the same geographic area who receive their care in the FFS sector. These adjustments are designed to modify HMO payments for expected variations in medical costs. However, this risk-adjustment is inadequate because it does not specifically adjust for the health status of enrollees (Lubitz, 1987; Epstein en Cumella, 1988; Ash et al. 1989; Newhouse et al., 1989). The current AAPCC risk groups are too heterogeneous which provides incentives for risk selection [5].

In principle there are two ways to prevent risk selection or cream skimming: refining the capitation payment formula and implementing additional procompetitive regulation (Van de Ven and Van Vliet, 1992). These strategies to eliminate insurers' incentives to risk select do not exclude each other. They can be used as complementary mechanisms (Swartz, 1995). The focus of this study is on refining the capitation payment formula. The results of this study are not only relevant for situations where competing health insurers or health plans are capitated like the sickness funds in the Netherlands or the HMOs in the United States, but also when providers are capitated, as in the UK [6].

In section 2.2 the requirements for ideal risk-adjusters are described. The next sections deal with potential risk-adjusters. Section 2.3 describes

---

[5] The evidence from the literature suggests that HMOs are subject to some favourable selection of new enrollees according to prior health care use and costs measures, although no selection bias at all is also a common result in many studies (Luft and Miller, 1988; Riley et al., 1989; Lichtenstein et al., 1991; Hellinger, 1995; Dowd et al., 1996). By enroling healthier individuals, HMOs need deliver less health care but they are compensated as if they enroled a costlier clientele. In that case Medicare has paid HMOs more than it would have paid for the same patients' care by fee-for-services providers (Brown et al., 1993; GAO, 1994b).

[6] As part of recent National Health Service (NHS) reforms in Britain, general practitioners (GPs) receive budgets to enable them to purchase selected hospital services on behalf of their patients (Glennester and Matsaganis, 1993). Adequate risk adjustment of these GP fundholding budgets is critical to the success of this part of the NHS reforms (Matsaganis and Glennester, 1994; Sheldon et al., 1994).

measures of prior use and prior costs; section 2.4 the use of diagnostic information from prior health care utilization; section 2.5 with self-reported health indicators and section 2.6 with other risk-adjusters proposed in the literature.

2.2 Requirements for ideal risk-adjusters

The most commonly used risk-adjusters are demographic variables like age and sex. The use of such crude risk-adjusters creates heterogeneous risk groups. By refining the capitation payment formula, i.e. extending the capitation formula with relevant risk-adjusters, the risk groups become more homogeneous. With more homogeneous risk groups, it is more difficult and therefore more costly to determine who are the low risk individuals within each risk group and the process of cream skimming would have to be more sophisticated, which makes it more expensive, while on average the profits of cream skimming will decrease (Van de Ven and Van Vliet, 1992). Therefore, extending the (demographic) capitation formula with relevant risk-adjusters on balance lowers the financial attractiveness of risk selection for insurers. However, the question is which are the relevant risk-adjusters for extending the capitation formula.

Ideally the risk-adjusters used to refine the capitation formula should meet the following requirements (Newhouse, 1986; Epstein and Cumella, 1988; Gerritse and Poelert, 1991; Van de Ven and Van Vliet, 1992; GAO, 1994a):

- *validity*: they should measure the need for health services utilization and define a system of adjustment in which the cells are relatively homogeneous with regard to this need for health care;
- *obtainability*: they should be obtainable for all potential enrollees without undue expenditure of time or money and without making the administrative system unfeasible;
- invulnerability to *manipulation*: they should not be subject to manipulation by insurers, providers, or the insureds;

- no perverse *incentives*: they should not provide incentives for inefficiency or for low quality care;
- finally, they should not conflict with the right to *privacy* of the insureds and health care providers.

It probably will be necessary to weigh these desirable properties one against another. There is considerable controversy over which of these properties is more critical.

Demographic factors are relatively poor predictors of future health care utilization of individuals. At a group level, the variation in expenditures can be explained better than at an individual level by a simple demographic and employee characteristics model (Robinson et al., 1991; Luft, 1995). In the present study the individual is the unit of measurement. Using data of individuals provides the opportunity to examine whether a particular capitation model is able to predict accurately the expenditures for subgroups identified on the basis of characteristics that are not included in the capitation model.

The predictive ability of a risk-adjuster depends on the kind of costs to be predicted and on the persons whose costs are predicted. The costs for catastrophic risks (like several forms of expensive long-term care) are much more predictable than the costs for noncatastrophic risks (like hospital care and physician services) (Van Barneveld et al., 1997). Outpatient services expenditures are generally more predictable than inpatient services (Wouters, 1991; Newhouse et al., 1993). Expenditures of special groups of chronically ill persons are much more predictable than the expenditures of the general population. Therefore, some studies focus on the development of risk-adjusted capitation systems for these specific groups like persons with end stage renal disease (Farley et al., 1996), disabled persons (Kronick et al., 1995, 1996) and persons with AIDS (Kahn et al., 1995). The purpose of the present study is to improve the demographic capitation system for the costs for noncatastrophic risks, both including outpatient and inpatient services, in a general population. This study focuses on prospective models that predict future, i.e. next year's, health care expenditures and not on concurrent or retrospective models that explain health care expenditures in the base-line period.

Demographic factors, the most commonly used risk-adjusters, are easy to obtain, invulnerable to manipulation and have no perverse incentives. However, they are poor predictors of future health care utilization of individuals. Demographic factors can predict 1 to 3 percent of the variation in total annual health care expenditures between individuals, where the maximum predictable portion of medical expenditure variation is estimated at about 20% (Newhouse et al., 1989; Van Vliet, 1992). Proposed health status risk-adjusters for refining the capitation formula are measures of prior costs and prior use (whether or not in combination with diagnostic information), perceived health status and functional health status (Lubitz, 1987; Epstein and Cumella, 1988). The next sections give an overview of the recent literature on these potential risk-adjusters.

2.3 Prior utilization

Prior utilization is considered to be an indirect measure of health status. Several measures of prior utilization have been used to predict health care expenditures, like total costs, costs for inpatient services, cost for outpatient services, any hospitalization, number of hospitalizations, number of days in the hospital, number of emergency room visits and the number of prior physician visits (Epstein and Cumella, 1988). Prior utilization appears to be the best single predictor of an individual's future health expenditures (Van de Ven and Van Vliet, 1992).

Thomas and Lichtenstein (1986) extended the set of AAPCC risk-adjusters with prior year's standardized payments in a study among the elderly, resulting in an increase of R^2-value from 0.3% to 5.9%. With prior utilization in the form of the number of inpatient admissions, number of (Medicare) part B claims and number of emergency room visits, 7.2% of variation in health care expenditures was explained. In a study among non-elderly adults Newhouse et al. (1989) showed that extending the set of AAPCC risk-adjusters with measures of prior year's inpatient and outpatient expenses increased the R^2-value from 1.6% to 6.4%, which was

estimated to be 44% of the maximum predictable variation. In a later study among children prior year's inpatient and outpatient expenses, in addition to a demographic capitation model, explained 20.7% of the variation in outpatient expenditures. This was about 57% of the maximum predictable variation (Newhouse et al., 1993). Schauffler et al. (1992) found an increase in R^2-value from 2.5% to 7.7% when they extended the AAPCC with two dummy variables for hospitalizations and physician visits during the last two years. Extending a demographic capitation model with prior year's total costs explained 7.2% of the variation in health care expenditures among Dutch privately insured persons of all ages. When prior year's outpatient costs (four variables for different types of outpatient costs) and inpatient costs were used instead of total costs the R^2-value increased from 2.4% for the demographic model to 7.3% for the prior utilization model (Van Vliet and Van de Ven, 1992).

In most studies prior use models are used to predict next year's expenditures. High and low use will regress towards mean use over time. Beebe (1988) found for groups biased on prior reimbursement and prior utilization a sharp regression toward the mean in the first year following the base year and a slow regression thereafter. The biased groups did not regress entirely to the mean even after six years. The most extreme groups in the base year remained farthest from the mean, which suggests that prior utilization captures some element of chronic health care need.

Prior utilization as a risk-adjuster for capitation payments has been criticized for two reasons (McClure, 1984; Porell and Turner, 1990). First, some differences in prior use among individuals could reflect differences in physician discretionary practice patterns. Capitation payments based on prior utilization would pay insurers in direct proportion to the prior utilization of their insureds without regard to the appropriateness of the care. Second, the payments would be based on an average relationship between prior costs and subsequent medical expenditures. Expected future costs, however, may differ widely for persons with high prior use associated with chronic medical conditions in contrast to those with acute illness. This might lead to perverse provider incentives or to new selection problems. Therefore, although an individual's prior costs is the best single

predictor of his future costs, it is not an ideal risk-adjuster in a capitation payment model.

A risk-adjuster should be a characteristic of the patient's expected need for services independent of his choice of provider. With the inclusion of a direct measure of prior utilization in a capitation model there is always the danger of rewarding inefficient providers and encouraging more utilization than is strictly necessary. To solve these problems, prior utilization risk-adjustments should be based on diagnoses associated with high medical expenses and judged to be nondiscretionary (Lubitz, 1987).

2.4 Diagnostic information from prior utilization

Various studies investigated the possibility of extending the capitation model with diagnostic information from previous utilization of either inpatient or outpatient services (Ash et al., 1989; Anderson, Steinberg, Powe et al., 1990; Weiner et al., 1991). The quintessence of such models lies in the allocation of people to a restricted number of groups according to the diseases diagnosed when they had contact with the health care system during the last year and incorporating this information in the form of dummy variables in the capitation model. The predictive accuracy of these models in terms of the percentage of predicted variance of per-person expenditures in the next year, was substantially higher than that of models with only demographic predictors. The next section describes capitation models that used diagnoses from prior utilization of inpatient care as a risk-adjuster and is followed by a section about the use of diagnoses from outpatient services and the use of drug claims to identify persons with chronic conditions to be used as a risk-adjuster.

2.4.1 *Inpatient care*

People with serious chronic diseases have a relatively high probability of being hospitalized in a given period, thus hospitalization could be used as an indicator of poor health. Anderson, Steinberg, Whittle et al. (1990) found among elderly persons discharged alive from the hospital for the 674 most common discharge diagnoses, significant variations by diagnosis code for the probability of dying, the probability of a readmission and mean level of future hospital expenditures. Anderson, Steinberg, Powe et al. (1990) estimated that nearly 50% of admissions are for temporary health problems. Thus, admissions should be classified to distinguish to what extent the associated diagnoses are indicative of chronic health problems and warrant higher future capitation payments.

In the literature several classification systems have been suggested for capturing the future costs of inpatient morbidity. Ash et al. (1986, 1989) constructed Diagnostic Cost Groups (DCGs), which can be seen as a refinement of the High Costs Next year (HCN) system, an a priori classification of hospital diagnoses according to a clinical judgement regarding whether they were likely to lead to High Costs Next year. Ellis and Ash (1988, 1995) and Ellis et al. (1996) further refined the Diagnostic Cost Group model. Anderson, Steinberg, Powe et al. (1990) developed the Payment Amount for Capitated Systems (PACS), which is based in part on the major diagnostic categories used in the prospective payment system for hospitals in the US. Both the DCG and the PACS model were developed with, and applied to, data from the US Medicare program, which mainly covers the elderly.

Ash and several coworkers developed various variants of DCGs, using both clinical and economic criteria. Ash et al. (1986, 1989) first classified diagnoses into 78 clinically homogeneous groups. These were further clustered into 9 groups according to the empirically determined similarities in the future costs of individuals hospitalized with different diagnoses. Some diagnoses were downgraded to group 0 –containing people who are not hospitalized– because the decision to hospitalize many individuals with the conditions concerned, was thought to be highly discretionary (e.g.

pneumonia and influenza). Persons with multiple hospital admissions in a year were uniquely assigned to the most expensive DCG to which any of their diagnoses belong. In the final capitation model the number of groups was reduced from 10 to 5 on the basis of regression results. This model gave an increase in R^2 of about 4 percentage points in comparison with a model that included only age, sex and welfare status (0.5% versus 4.7%). The costs in the following year of those individuals who had a non-discretionary hospital admission appeared to be three times as high as the costs of those without admission (or with a discretionary admission). This difference appeared to persist over a period of at least three years. Gruenberg et al. (1989) showed that the same group of individuals experienced significantly higher hospital use rates over a period of at least six years. These results suggest that the DCG classification captures some element of long-term chronicity.

Ellis and Ash (1988, 1995) took a somewhat different approach. First, clinical judgment was used to remove diagnoses for which the decision to hospitalize may involve high levels of discretion. The original 78 clinically homogeneous groups were subdivided to reflect both discretion ratings and the ability to further differentiate future costs among persons hospitalized with particular diagnoses. For instance, cancers that were previously split into seven clinically homogeneous groups, now were divided into 24 separate groups. 104 separate so-called DXGROUPS were formed based on clinical judgement. Next, the same clustering procedure was followed as before, this time resulting in the formation of 8 groups based on the distribution of future expenditures. In the capitation model the diagnostic group 0 for the diagnoses with lowest future expenses was expanded with people who had not been hospitalized and those who were hospitalized with a high discretion diagnosis. The reduction in R^2 due to the classification of high discretion diagnoses in the lowest group was 1.4 percentage points (from 5.2% to 3.8%).

In later work the DCG model was further extended (Ellis et al., 1996). A first extension added secondary diagnoses from inpatient bills, diagnoses from hospital outpatient claims and diagnoses from bills for ambulatory or inpatient physician services to principal hospital inpatient diagnoses. Using

Medicare claims data Ellis et al. showed that the predictive accuracy, in terms of R^2-value of a demographic model with the AAPCC risk-adjusters improved by 4.5 percentage points to 5.5% when DCGs based on an individual's single highest (future) costs principal inpatient diagnosis, were added to the capitation formula. When secondary inpatient, hospital outpatient and physician inpatient and outpatient diagnoses were added to the principal inpatient diagnosis and persons were classified into a DCG based on their single highest costs diagnosis the R^2 increased modestly with 0.8 percentage point.

DCGs are mutually exclusive and exhaustive. When multiple chronic health problems exist, a single diagnosis can describe a person's health status only partially. Therefore, the second extension expanded the risk-adjustment framework to account for multiple medical conditions that people may experience. This was called the hierarchical coexisting conditions (HCC) model. First, coexisting condition groups were formed by combining clinically homogeneous DXGROUPS belonging to a major body system or disease type by costliness and clinical relation. Next, hierarchies were created among subsets of the coexisting conditions based on clinical judgement. The hierarchies specified that a person with multiple, clinically related coexisting conditions was assigned only to the highest ranked among these related coexisting conditions. For example, a person in the HCC 'metastatic cancers' was not allowed to be in any of the other six HCCs in the neoplasm hierarchy. This resulted in 66 Hierarchical Coexisting Conditions (HCC) diagnostic categories. Thirty-two of these HCC groups were excluded from the final model because they were not predictive of significantly higher Medicare expenditures, medically ambiguous, had relatively ambiguous criteria for coding on claims or were difficult to audit or verify. The HCC model which contains dummy variables for 34 HCC plus age and sex, explained 8.1% of the variation in medical expenditures. Extending the HCC model with dummy variables for 11 life sustaining medical procedures gave a small improvement in R^2 of 0.6 percentage points. A further extension with three groups of principal inpatients diagnoses resulted in an improvement of 0.3 percentage points.

The Payment Amount for Capitated Systems (PACS) of Anderson, Steinberg, Powe et al. (1990) is, apart from age, sex and disability status, based on the major diagnostic category (MDC) associated with each hospitalization in previous years, the chronicity of each clinical disorder that resulted in an admission (chronic, acute with possible sequelae or acute self-limited), the number of admissions and ambulatory resource use. There seem to be three important differences with the studies by Ash et al. (1986, 1989) and Ellis and Ash (1988): (a) the use of the number of hospitalizations; (b) the inclusion of a dummy variable for outpatient expenses; and (c) the diagnostic categories are based solely on clinical criteria and not on economic criteria. In view of these differences it is not surprising that Anderson, Steinberg, Powe et al. found that the predictive accuracy in terms of R^2 for the PACS model was approximately 8 percentage points higher than that of the discretion DCG model from Ellis and Ash (1988), namely around 13% versus about 5% [7].

Van Vliet and Van de Ven (1993) applied the original DCG model of Ash et al. (1986), the discretion DCG model of Ellis and Ash (1988) and the PACS model to Dutch data of privately insured persons of all ages. Extending a demographic capitation model with dummy variables for the nine original DCGs increased the R^2 from 3.5% to nearly 6.7%. The discretion DCG model explained 5.7% of the variation in health care expenditures. The PACS model yielded the highest R^2 with 8.4%. The difference in R^2 between the PACS and the DCG models was smaller for this population of both aged and non-aged persons than for the Medicare beneficiaries in the study of Anderson, Steinberg, Powe et al. (1990). Extending the original DCG and PACS model with prior year's costs further improved the R^2 to about 12%.

Models incorporating diagnostic information from prior hospitalizations are better able to predict future medical expenditures than a demographic capitation model. These diagnostic models create incentives for health plans to enrol and appropriately treat high costs individuals. Diagnosis-based

[7] In the study of Anderson, Steinberg, Powe et al. (1990) persons whose expenditures were in the top 1% of the distribution were excluded.

risk-adjustment models can also create perverse incentives, i.e. inappropriately hospitalizing people in order to obtain a higher capitation payment for them in the future. Assuming that no individuals would be hospitalized without any medical problem, the extent of this problem depends on the present marginal costs and the future marginal benefits of treating individuals on either an inpatient or outpatient basis. This problem could be reduced by a good monitoring system and by not rewarding admissions for high discretion diagnoses (Van Vliet and Van de Ven, 1993). Using diagnoses to adjust capitation rates has negative aspects similar to concerns that have arisen regarding the use of diagnoses in Medicare's hospital prospective payment system in the United States. The administrative burden for HMOs would increase. Diagnoses are subject to manipulation by the provider. An HMO equivalent of a diagnosis-related group (DRG) creep could emerge as providers attempt to assign more patients to categories with higher reimbursement rates (Manton et al, 1989). Hsia et al. (1988) showed that DRG creep occurred, resulting in overpayment to hospitals for patients covered by Medicare. Ideally, capitation payments should be relatively insensitive to variations in coding practices and to treatment choices such as rates of hospitalization (Ellis et al., 1996).
In a situation where diagnostic information is recorded on a regular basis, a capitation system (partly) based on these diagnoses seems administrative feasible. Nelson and Arnold (1990) concluded from the assessment of the diagnostic cost group pilot demonstration that 'the DCG payment system is operationally feasible, but more complex than the AAPCC system. It imposes additional responsibilities on both HCFA and the HMOs, and its successful operation in the future will require more intensive monitoring'.

Extending a demographic model with a measure of inpatient morbidity improves the predictive accuracy of the capitation model. However, even in an elderly population a small fraction only will be hospitalized in a year. Not all individuals suffering from chronic conditions with related predictable expenses are admitted to a hospital in any given year. Some have argued that this limits the utility of inpatient morbidity models, especially for non-aged populations (Hornbrook et al., 1991). Models using

outpatient morbidity whether or not in combination with inpatient morbidity could be an alternative.

2.4.2 *Outpatient care*

The Ambulatory Care Group (ACG) system is a case-mix measure useful in predicting the utilization of ambulatory health services (Starfield et al., 1991; Weiner et al., 1991). The building blocks of the ACG system are 34 Ambulatory Diagnostic Groups (ADGs). ADGs are clusters of ambulatory diagnostic codes. The clustering was based on expected resource use. The ACG system was formed by 51 mutually exclusive ACGs constructed from each individual's unique combination of ADGs, along with age and gender. For a non-aged population of HMO beneficiaries age and sex explained 5% of the variation in next year's ambulatory visits and 3% of the variation in next year's ambulatory charges between beneficiaries. Extending the demographic model with dummies for ADGs increased the R^2-values to 23% respectively 21%. With the ACG system 20% of the variation in visits and 18% of the variation in ambulatory charges was explained.

In a later study by the Physician Payment Review Commission the ADGs and the ACG system were used to predict both outpatient and inpatient charges for persons of all ages. In this study age and sex explained 1.6% of the variations in total charges. The R^2 for the model with ACGs was 3.1%. The model with age, sex and ADGs yielded the highest R^2-value with 5.2% (PPRC, 1994b). This results show that although the ADGs and ACGs were developed as a case-mix measure to be used in an ambulatory setting, they also have some predictive ability for both outpatient and inpatient charges.

Researchers from the Johns Hopkins University developed two new diagnosis-oriented methodologies for setting risk-adjusted capitation rates for managed care plans contracting with Medicare by combining components of the ACG model with some components of the PACS model (Weiner et al., 1996). Both Medicare risk-adjuster models incorporated ADGs, the basic morbidity classification of ACGs. ADGs

that were relatively poor predictors of future medical costs in the elderly population were excluded. The models derived from PACS use three demographic risk assessors (age, sex and prior disability status) and an inpatient measure based on the Major Diagnostic Category (MDC). The first model, the ADG-MDC model, in addition to containing demographic variables also included dummies for 13 selected ADGs and count variables for 15 selected MDCs. The assignment to ADGs was based on diagnoses (either primary or secondary) noted by providers during face-to-face encounters in the ambulatory visits setting. The second model, the ADG-Hosdom model, contained demographic variables, dummies for 13 selected ADG (based on all available diagnoses on inpatient and outpatient claims and those noted by clinicians during face-to-face encounters in both the ambulatory and inpatient setting) and a 'Hospital Dominant' (Hosdom) marker. The Hosdom marker is a binary variable indicating the presence within an individuals claims records of one or more of the 843 ICD-9-CM codes that were serious enough to usually be treated on an inpatient basis. For the 843 diagnoses at least 50 percent of the patients had been hospitalized. The ADG-MDC model explained 6.3% of the variation in next year's medical costs between Medicare beneficiaries; the ADG-Hosdom explained 5.5% of the variation. The demographic AAPCC model yielded a R^2 of only 1.0%. Weiner et al. showed that by incorporating diagnostic information the ADG-MDC and the ADG-Hosdom models were better able to predict future medical expenditures than a demographic capitation model, for both randomly selected and non-randomly selected individuals.

Hornbrook et al. (1991) used diagnoses recorded in outpatient medical records and prescribed drugs to predict future total annual expenses for medical care in an employed population. The diagnoses were aggregated into 16 morbidity classes. Based on clinical judgement, diagnoses were classified according to the nature of the disease and intensity of the expected response from the medical care system. Drugs were classified into 27 therapeutic classes. Per class the number of drugs orders was counted. This measure captured the general therapeutic action desired by the physician from which could be inferred the nature of the disease under treatment. Counts of drugs orders by therapeutic class also assessed the

variety of drugs prescribed for an individual and the overall intensity of prescribing activity. Age and sex explained 2.1% in variations in next year's total expenses. Extending this model with morbidity classes increased the R^2 with 4 percentage points; an extension with drugs classes with almost 3 percentage points. A model with age and sex, morbidity classes and drugs classes yielded the highest R^2 of 6.9%.

Von Korff et al. (1992) used population based automated outpatient pharmacy data to construct a measure of chronic disease status. They developed the Chronic Disease Score (CDS) relying on clinical judgement. The score was based on the number of different diseases, complexity of the treatment regimen, whether the disease was progressive or not, and disease rather than symptom management. The CDS was correlated with physician-rated disease severity and it was found to predict hospitalization and mortality in the following year after controlling for age, gender and health care visits. Von Korff and colleagues concluded that scoring automated pharmacy data can provide a stable measure of chronic disease. Johnson et al. (1994) replicated the study and found that the CDS was stable from year to year and had construct and predictive validity.

Clark et al. (1995) developed a revised version of the CDS, covering a wider range of medications than the original CDS. Rather than relying on physician judgement of disease severity to assign CDS weights, weights for individual drug classes were estimated empirically for the revised CDS. The CDS is a set of dummy variables that indicate a pharmacy prescription during a six month period for a medication or medication class representing particular chronic diseases. 28 different conditions were distinguished. The revised CDS model containing the 28 binary variables forming the revised CDS together with age and sex explained 10% of the variations in total medical expenditures and 23% of the variations in outpatient costs of adults (18 years and older) enroled in an HMO in the next six months period. Age and sex alone explained 3% and 6% of the variations in total charges respectively outpatient charges. Clark et al. estimated also an Ambulatory Diagnostic Group (ADG) model containing 34 dummy variables for ADGs and age and sex. The ADG model yielded R^2-values comparable with the revised CDS model: 8% for total costs and

21% for outpatient costs. The combination of revised CDS and ADGs showed only marginally greater predictive power than either one alone. This suggest that the drug information used in the revised CDS and the ambulatory diagnoses from the ADGs captures the same part of the predictable variations in future health care expenditures. The revised CDS and the ADGs can replace each other.

Both diagnostic information from outpatient health care visits and drugs prescribed for chronic conditions have predictive ability for future health care expenditures. The disadvantages of perverse incentives and possibility of manipulation as discussed for inpatient morbidity models also apply to models using ambulatory diagnoses. Physicians can readily initiate visits by providing appointments for follow-up. A system based on diagnosis might be gamed by encouraging the recording of more diagnoses, thus leading to classification of patients into higher reimbursement categories (Starfield et al., 1991).
Prescribed drugs work on a principle similar to morbidities. They capture underlying disease validated by a doctor's drug order. When used as a risk-adjuster only prescriptions of drugs to treat chronic conditions, as in the CDS, should be rewarded with a higher capitation payment next year. Using the CDS as a risk-adjuster creates perverse incentives, i.e. encouraging inappropriately prescribing drugs in order to obtain a higher capitation payment. There may be a possibility of manipulation, because knowledge of weights may lead providers to alter prescription behavior to maximize payment (Clark et al., 1995).
The administrative feasibility of a capitation system based on either ambulatory diagnoses or prescribed drugs depends on the availability of data. In a situation were the information is routinely recorded and automated such capitation systems seem feasible.

2.5 Self-reported health indicators

Self-reported health indicators, mostly obtained by surveys, like perceived health status, functional health status and self-reported chronic conditions are predictors of future health care expenditures (Thomas and Lichtenstein, 1986; Van Vliet and Van de Ven, 1992; Hornbrook and Goodman, 1996). Perceived health status, also called subjective health, is sometimes measured with a single item or with multi-item questionnaires. Chronic conditions refer to the answers to questions like: 'Has a doctor ever told you that you have ...'. Functional health status and disability status are often measured by the same kind of questions like limitations in activities of daily living (ADL) and instrumental activities of daily living (IADL). Lichtenstein and Thomas (1987) showed that for the Medicare population (IADL) functional health status remained stable over time.

Thomas and Lichtenstein (1986) extended the set of AAPCC risk-adjusters with several self-report health indicators in a study among the elderly. The improvement in R^2 was about 1 percentage point for chronic conditions and ADL; 2.3 percentage points for perceived health (single item or multi-item scale). Extending the AAPCC with a scale measuring IADL, whether or not in combination with an ADL-scale, resulted in the greatest improvement in R^2 from 0.3% to 3.9%. Schauffler et al. (1992) found an increase in R^2-value from 2.5% to 4.8% when they extended the AAPCC with disability status measured by 18 questions on physical functioning. In a study among Dutch privately insured persons of all ages, a demographic capitation model including age, sex, region and insurance coverage explained 2.8% of the variations in next year's total expenditures. Extending this demographic model with chronic conditions increased the R^2 to 7.1%. A further extension of this model with physical impairments resulted in an R^2-value of 7.7%. The most comprehensive model, which also included self-rated general health status, explained 10.9% of the variations in medical expenditures (Van Vliet and Van de Ven, 1992).

In a study by the Physician Payment Review Commission among persons of all ages, age and sex explained 1.6% of the variations in total charges.

Age, sex and a single general health item explained 3% of the variations in medical expenses. Extending the demographic model with scores on the 8 scales of the Short-Form (SF) 36 health survey improved the R^2 with 3.3 percentage points. This is similar to the improvement of 3.2 percentage points by adding chronic conditions. A combination of SF-36 health survey scales and chronic conditions increased the R^2 to 6.2%. The model containing age, sex, chronic conditions, SF-36 health survey scales and one-year ADGs yielded the highest R^2-value of 7.0% (PPRC, 1994b).

Hornbrook and Goodman (1995) explored the risk structure of employed HMO members using the RAND-36 health survey. This health survey comprises nine scales measuring different dimensions of health. The nine RAND-36 scales explained 3.9% of the variations in total charges between members. When age and gender were added to the model the R^2 increased to 4.7%. A model with age, gender and five [8] instead of nine RAND-36 scales yielded the same R^2. When the model was extended with interactions between age, gender and the RAND-36 scales the R^2 further increased to 6.5%. Hornbrook and Goodman concluded that self-reported perceived/functional health status provided substantially better predictive performance than demographic factors, but demographic factors were required to achieve unbiased forecasts. In another study Hornbrook and Goodman (1996) used demographic factors, functional health status (the RAND-36 health survey) and self-reported chronic conditions to predict future medical expenses of employed HMO members. Chronic conditions explained 2.3% of annual per capita expenses; chronic conditions and demographic factors 3.1%. A model comprising demographic factors, the nine RAND-36 health survey scales and chronic conditions increased the R^2-value to 5.5%. They also found evidence of interaction effects between functional/perceived health status scales and disease classes.

Gruenberg et al. (1996) used health status measures from the Medicare current beneficiary survey to improve the predictive accuracy of the AAPCC. Their results for a Medicare population of the elderly were

[8] The five scales are: physical functioning, social functioning, limitations in role function caused by physical problems, perceived change in health and general health perceptions.

comparable with the results of Hornbrook and Goodman (1996) for a non-aged population of HMO members. The demographic model explained only 0.7% of the variation in medical expenses of Medicare beneficiaries. Extending the demographic model with self-reported health status (single item) and self-reported chronic conditions improved the predictive accuracy in terms of R^2-value to 4.1%. Extending the demographic model with information on limitations in ADL increased the R^2 to 3.2%; a further extension of this ADL model with a measure of IADL/disability increased the R^2 to 4.9%. A comprehensive model comprising demographic factors, self-reported health status, chronic conditions, ADL and IADL/disability explained 6.0% of the annual per capita medical expenses. For the sake of comparison they also included a prior use model in their study. The prior use model was the comprehensive model extended with three prior use variables, part B, home health visits and the number of inpatient days. This prior use model yielded the highest R^2 of 13.4%.

Hornbrook and Goodman (1995) pointed out that self-reported health status offer some important advantages. The self-report is the most appropriate method of measuring perceived well-being and attitudes about health and medical care that govern demand for care. The self-report can provide data where non exist under other approaches. Morbidity based models, like ACGs or DCGs, fail to account for unmet needs. Standardized surveys can obtain consistent data across health plans. Using self-report health surveys for risk-adjustment has some disadvantages: it may be administratively infeasible, administration costs are high and there is a possibility of gaming. One can question whether mailed health surveys are appropriate for detecting and assessing the health status of vulnerable populations. Another unresolved question is how to handle nonresponse in a risk-assessment model.

2.6 Other risk-adjusters

Other risk factors that could be used as risk-adjusters for capitation payments are sociodemographic or socioeconomic factors, behavioral risk factors, mortality and physiological risk factors. Sociodemographic factors are characteristics like marital status, family size and indicators of social economic status are occupation, income and education. These factors hardly improved the predictive accuracy of capitation models already including age and sex (Epstein and Cumella, 1988; Van Vliet and Van de Ven, 1992; PPRC, 1994b).

Behavioral risk factors such as smoking, drinking alcohol, lack of physical activity and obesity were examined in the study mentioned earlier by the Physician Payment Review Commission (PPRC, 1994b). Models were tested that extended the demographic (age, gender) model with each of the four behavioral risk factors separately as well as combined. The R^2 for the model that included smoking and demographics and one that included body mass index and demographics were below that of the demographic reference model. The model with demographics and all four behavioral risk factors improved the R^2-value with 0.9 percentage points to 2.5%. It was concluded that behavioral risk factors did not perform well in predicting health care charges at the individual level.

Tolley and Manton (1984) proposed to use cost-weighted disease-specific mortality as a risk-adjuster. The problem with mortality or the probability of dying as a risk-adjuster is to identify persons at high risk of death (Newhouse, 1986). Culler et al. (1995) showed that in a study among elderly patients who died between 1984 and 1991, the concentration of resources consumed in the last year of a respondent's life was only marginally significant in explaining total real hospital charges over an 8-year observation window. The most important explanatory factors were variables used to control for the distribution of comorbidity, population density and a variable indicating whether or not the patient's death was related to an acute myocardial infarction. Nooren and Van Vliet (1994) concluded from an empirical analysis of data of Dutch privately insured

persons of all ages that -next to age and sex- mortality would improve capitation payments at best marginally.

Howland et al. (1987) showed that physiological measures (respiratory function, blood sugar, blood pressure, cardiac function, smoking and weight) and prior hospitalizations were better predictors of hospitalizations in a two-year follow-up period than demographic factors for a group of 60-65 years old participants in the Framingham Heart Study. The contributions of the physiological risk factors and prior hospitalizations were about equal and independent. In a later study among Medicare beneficiaries from the Framingham Heart Study cohort Schauffler et al. (1992) used chronic disease risk factors to predict next year's per capita medical expenses. The risk factors smoking, blood pressure, serum cholesterol, blood sugar, respiratory function and two-year probability of cardiovascular disease explained 6.5% of the variations in medical expenses; together with the AAPCC risk-adjusters the risk factors explained 6.6% of the variation in per capita expenses. Extending this model with prior hospitalizations and prior physician visits improved the R^2 with 4.3 percentage points to 10.8%. A model including the AAPCC risk-adjusters, the five above mentioned risk factors, the prior utilization measures and a survey measure of disability status yielded the highest R^2 of 13.0%.

In a study among non-elderly adults Newhouse et al. (1989) used several measures of physiological health to extend the set of AAPCC risk-adjusters. Extending the AAPCC model with dichotomous variables for physiological health increased the R^2-value from 1.6% to 3.8%; with continuous variables for physiological health to 4.2% and with dichotomous variables based on claims to 4.5%. A further extension of these models with prior year's inpatient and outpatient expenses resulted in R^2-values of 8.0% and 8.7%, which was estimated to be 55% respectively 60% of the maximum predictable variation. A set of measures on functional status, self-rated general health perceptions, mental health and a variety of self-reported chronic conditions could hardly further improve the predictive accuracy of these models. A later study among children gave similar results. Dichotomous and continuous physiological health, in addition to a demographic capitation model, explained 9.9% respectively

10.5% of the variation in outpatient expenditures. Extending these models with prior inpatient and outpatient expenses increased the R^2-value to 23.5%. This was about 64% of the maximum predictable variation (Newhouse et al., 1993).

The major advantages of using physiological risk factors as a health status risk-adjuster for capitation payments, are their predictive accuracy and their strong association with chronic disease. Another advantage of physiological measures is that they can be objectively measured and verified, reducing concerns of manipulation in reporting risk levels and misclassification of risk status. As one relies on physiological or biochemical measures of risk, in an effort to reduce potential manipulation, the collection of risk factor data becomes more invasive and expensive. The major disadvantages of using physiological risk factors as risk-adjusters is the additional administrative burden and costs associated with collection and periodic assessment of risk factors (Schauffler et al., 1992). Risk factor data will not necessary reflect the relationship of increasing expenditures with less healthy values of risk factors. This is especially the case if treatment alters the physiological measure and less healthy persons utilize more resources (Newhouse et al., 1989). For example, a person may have high blood pressure which is under control and below 160/90. This person is considered low risk. However, the medical expenditures to maintain control are expected to be much higher for this person than for the person whose blood pressure is naturally low.

2.7 Conclusion

Risk-adjusters for capitation payments should have properties like *validity*, *invulnerability to manipulation*, *obtainability* and provide *no perverse incentives*. A valid risk-adjuster should measure the need for medical care and relate to health status. Risk-adjusted capitation payments should account for predictable variations in annual per-person health care expenditures, as far as these are related to health status. Demographic

factors, the most commonly used risk-adjusters, are relatively poor predictors of future health care utilization of individuals. Risk-adjustment based only on demographic factors is inadequate because it does not adjust sufficiently for health status. Therefore, measures of prior utilization (whether or not in combination with diagnostic information), self-reported health indicators like perceived and functional health status and physiological measures has been proposed for refining capitation payment formulae.

Prior utilization appears to be the best single *predictor* of an individuals *future health expenditures*. Prior costs models as well as prior use models with measures like any hospitalization, number of hospitalizations, number of days in the hospital, number of emergency room visits and the number of prior physician visits have predictive ability for future health care expenditures. Models using diagnostic information from either inpatient or outpatient health care visits and drugs prescribed for chronic conditions have predictive accuracy too. Models incorporating this information are better able to predict future medical expenses than a demographic capitation model. Incorporating one or more survey-based health indicators like perceived health status, functional health status or disability status and self-reported chronic conditions in a demographic capitation model also improves the predictive accuracy of the model. The self-report is the most appropriate method of measuring perceived well-being and attitudes about health and medical care that govern demand for care. The self-report can provide data for unmet needs. Physiological measures like blood sugar, blood pressure, respiratory function, cardiac function, smoking and weight can improve the predictive accuracy of a demographic capitation model. The major advantages of using physiological risk factors as a risk-adjuster for capitation payments are their predictive accuracy and their strong association with chronic disease.

With the inclusion of a direct measure of prior costs or prior use in a capitation model there is always the danger of rewarding inefficient providers and encouraging more utilization than is strictly necessary. The extent of this problem of perverse *incentives* depends on a health plan's present marginal costs for inappropriate treatment and its future marginal

benefits. To reduce these problems various studies investigated the possibility of extending the capitation model with diagnostic information from previous utilization of either inpatient or outpatient services. These models create incentives for health plans to enrol and appropriately treat high costs individuals. Like direct measures of utilization they can also create perverse incentives, i.e. encouraging inappropriate utilization in order to obtain a higher capitation payment in the future. In the case of an inpatient morbidity model this means hospitalizing a person who could be treated just as well in an ambulatory setting. In case of an ambulatory morbidity model physicians can initiate visits, which are not strictly necessary, by providing appointments for follow-up. In case of prescribed drugs models one can think of prescribing drugs in a situation where, for instance, the advice to change one's life style would suffice. Assuming that no individual would be hospitalized without some medical problem, reduces the problem of perverse incentives for the inpatient morbidity models. However, many patients expect the outcome of a doctor's office visit to be a prescription.

With morbidity and prescribed drugs models there may be a possibility of *manipulation*, namely by 'inflating' diagnoses to move hospitalizations to better paid categories or in case of ambulatory morbidity models by recording more diagnoses, thus leading to classification of patients into higher reimbursement categories. This also holds for prescribed drugs models where knowledge of weights may lead providers to alter prescription behavior to maximize payment. Although there is a possibility of gaming, self-reported functional and perceived health status may be less subject to manipulation by health plans, because these variables measure enrollee perceptions, not diagnostic labels. Physiological risk factors can be objectively measured and verified, reducing concerns of manipulation in reporting risk levels and misclassification of risk status.

The use in practice of a capitation system based on either inpatient or ambulatory diagnoses or prescribed drugs depends on the availability of data. Uniform, generally accepted classification systems for diagnoses and drugs are necessary to record this information consistently across health plans. In a situation were the information is routinely recorded and

automated such capitation systems seems feasible. Using self-report health surveys for risk-adjustment has disadvantages with regard to *obtainability* and *feasibility*: it may be administratively infeasible and administration costs are high. The major disadvantages of using physiological risk factors as risk-adjusters is the additional administrative burden and costs associated with collection and periodic assessment of risk factors.

The potential health status risk-adjusters reviewed above all have some of the desirable properties a risk-adjuster ideally should have. The empirical analyses of this study will focus on diagnostic information, because this potential risk-adjuster, as Kronick et al. (1996) described it, 'seems to strike the best balance of practicality, accuracy and appropriate incentives'. Health survey information and information on physiological risk factors is not available for large groups of people. On the other hand, health plans often routinely record information on prior utilization and diagnoses. Although direct prior utilization measures are better predictors of future medical expenses, diagnoses-based models are preferred over prior utilization models because of their appropriate incentives.

At the start of this study most sickness funds in the Netherlands recorded diagnostic information from hospitalizations on a regular basis [9]. Therefore, in this study inpatient morbidity models will be examined. Among the models based solely on inpatient diagnostic information the PACS model had the highest R^2. However, since this appears to stem from the inclusion of variables like the number of hospitalizations, that are quite vulnerable to manipulation by health plans, we believe this not to be the most adequate model. Therefore, the DCG model developed by Ash and coworkers will be used as a starting point for the empirical analyses. The next chapter describes the data and methods used for these analyses.

[9] Ambulatory diagnoses are not routinely recorded by most Dutch sickness funds and are not available in the data set used for the empirical analyses in this study. Only recently information on prescribed drugs became available at the individual level.

3

Data and method

3.1 Data

The empirical analyses in this study are based on a panel data set with administrative data, which is complemented with data gathered by a mailed health survey. The health survey data are available for only a part of the persons in the panel data set. For these persons information from the panel data set can be matched with the health survey data. The next section describes the panel data and is followed by a section concerning the survey data.

3.1.1 *Panel data*

The panel data set contains administrative data from 'Zorg en Zekerheid', a sickness fund working in the western part of the Netherlands with about 420,000 members. The membership of this sickness fund is globally representative for all 9.7 million Dutch sickness fund members. The data set represents all 245,720 individuals that were continuously enroled with Zorg en Zekerheid during the four year period 1988-1991. For efficiency reasons not all of these were actually included in the data set. All individuals hospitalized in 1988 (about 19,000) are included in the data set, supplemented by a random sample of about 31,000 persons from the group of individuals not hospitalized in that year. The reason for this stratification is to get an over-representation of people with a relatively poor health status, which is the most interesting group in the context of capitation payments. All results presented are corrected for the stratification, by means of weighting for age/sex and whether or not hospitalized in 1988.

All individuals in the data set were continuously enroled during 1988-1991. Since 1992 persons in the panel data set can leave the sickness fund. Some people died and others changed sickness funds, resulting in an annual loss varying from 3.1% in 1992 to 2.6% in 1994. At the same time the data set is lacking information about young children, born in 1988 or later. To compensate for the loss of persons and to make the panel data set representative for all age groups, including young children, information for another 8,500 persons was included in the data set in 1994. All these individuals were continuously enroled during 1991-1993. Children born after 1990 were continuously enroled from date of birth until the first of January 1994.

Table 3.1 describes the study population on the first of January 1994. The table shows descriptive statistics for demographic variables for the study population, stratified according to the moment persons were included in the data set.

Table 3.1 shows that the newly added sickness fund members are younger than those who were initially included. Type of insurance indicates the compulsory cause for enrolment with the sickness fund, for example wage earners with a salary below a certain threshold, recipients of disability or unemployment benefits and the elderly with low incomes. Young children are assigned to the category 'others' in table 3.1. This explains the relative high percentage in this category for the individuals included in the second stage.

For each member initially included in the data set, administrative information on hospitalizations (when applicable) and annual health care expenditures are available for seven years, from 1988 to 1994. For the second stage members, the data set comprises the information for the period 1991-1994.

The annual per-person health care expenditures include the costs of inpatient room and board, both inpatient and outpatient specialist care, dental care, obstetrics and maternity care, paramedical services (physiotherapy and speech therapy) and sick-transport. The costs of drugs

Table 3.1 Descriptive statistics for demographic variables

	initially included	included in second stage	total
Sex:			
men	44.8 %	44.5 %	44.7 %
women	55.2 %	55.5 %	55.3 %
Degree of urbanization:			
very strongly urban/big city	11.4 %	12.5 %	11.6 %
strongly urban	26.8 %	28.8 %	27.2 %
moderately urban	25.7 %	25.4 %	25.7 %
little urban	21.7 %	21.7 %	21.7 %
rural	14.4 %	11.6 %	13.8 %
Type of insurance:			
employed policy holders	44.9 %	41.9 %	44.3 %
disabled policy holders	6.9 %	2.7 %	6.0 %
others	48.2 %	55.4 %	49.7 %
Age:			
0 years	--	5.7 %	1.2 %
1 - 5 years	--	24.9 %	5.3 %
6 - 15 years	10.6 %	5.1 %	9.4 %
16 - 25 years	13.7 %	27.6 %	16.7 %
26 - 35 years	22.2 %	16.4 %	21.0 %
36 - 45 years	15.8 %	7.1 %	13.9 %
46 - 55 years	12.9 %	5.3 %	11.3 %
56 - 65 years	10.4 %	3.9 %	9.0 %
66 - 75 years	8.1 %	2.3 %	6.9 %
≥ 76 years	6.3 %	1.7 %	5.3 %

prescribed by physicians were not available. The costs of care provided by the general practitioner (GP) are excluded because GP's receive a uniform annual fee for each sickness fund member in their practice regardless of medical consumption. All cost data refer to actual charges. From 1988 until 1991 not all costs were registered on an individual level. For the period 1988-1994 table 3.2 shows the mean costs per member according to the sickness fund administration and the mean costs per member in the panel

data set. In 1990 about 46% of the costs are missing. This is the year in which the sickness fund changed to another administrative system.

Table 3.2 Mean costs in Dutch guilders per member per year according to administration and in panel data set

	mean costs [a] administration	mean costs [a] in panel data set	panel data set / administration
1988	1035	808	78 %
1989	1037	680	66 %
1990	1116	605	54 %
1991	1251	979	78 %
1992	1310	1256	96 %
1993 [b]	1379	1338	97 %
1994 [b]	1437	1332	93 %

[a] Costs refer to total costs, exclusive costs for GP.
[b] In 1993 and 1994 the mean costs are calculated for both the initially included sickness fund members and the members included in the second stage.

In the period 1988 to 1993 the proportion of persons with one or more admissions to the hospital in a year ranges from 6.4% to 6.8%. In this six-year period 28.2% of the persons who were initially included in the data set, had at least one hospital admission. For each hospital admission in the period 1988-1993 the diagnosis is known in the form of the relevant code from the ICD-9-CM (International Classification of Diseases, ninth edition, Clinical Modification) coding system (SMR, 1980). In principle, the disease is recorded that is diagnosed on admission because when a member is hospitalized the sickness fund has to be notified of the reason for admission. However, notification is often delayed until after the discharge in which case the more informative discharge diagnosis is recorded. Physicians hospitalizing a patient make the diagnoses. These diagnoses are processed by the hospital which provide them to the sickness fund. It may happen that during one hospitalization several diagnoses are made. For instance when complications occur or when a specific disease is diagnosed for a person who was hospitalized for general symptoms. For each year

separately the diagnoses can be classified into DCGs. For people with more than one diagnosis, the diagnosis classified in the highest DCG is used. From 1990 onwards the diagnoses for some admissions are missing [10]. The persons concerned are classified to a separate DCG called 'DCG diagnosis unknown'.

3.1.2 *Health survey data*

In 1993 a mailed health survey was conducted. The main purpose of the survey was to gather information on health status and (additional) medical consumption. In February a first mailing, containing the questionnaire with the cover letter, was sent. After one week everyone received a postcard reminder, which served both as a 'thank you' for those who had already responded and as a friendly reminder for those who had not. Three weeks after this postcard reminder a letter with a new questionnaire was sent to the nonrespondents only. Two weeks after this, the nonrespondents received a final mailing. The design of this procedure was guided by Dillman's recommendations (Dillman, 1978).

About 15,000 people received the health survey. These persons formed a random sample of those who were initially included in the panel data set and did not receive institutional care. The health survey was sent to 13,472 adults between 15 and 90 years old and to the parents of 1509 children aged 5 to 14 years [11]. The parents were asked to complete the questionnaire for their child. Survey data were compared with the administrative data for date of birth and sex to make sure that the eligible person completed the questionnaire and not someone else in the household. For the adults 413 questionnaires were completed by another person; 64 parents completed the questionnaire for another child than was asked for. Questionnaires completed by someone other than the eligible person were considered as a nonresponse, resulting in a net response rate of

[10] This mainly concerns admissions to two hospitals.

[11] Age at January first, 1993.

70.0% for adults and of 75.4% for children. The net response rate for the total sample was 70.4%.

An analysis of the nonresponse showed that response was associated with age, sex, degree of urbanization and type of insurance. After correcting for differences in demographic characteristics respondents and nonrespondents differed in utilization of several types of care. Relatively more users than nonusers responded. Response was not associated with utilization of care related to severe conditions such as inpatient hospital care. The conclusion from the nonresponse analysis was that nonresponse bias or selection bias resulted in a small overestimation of utilization of outpatient care (an article addressing the nonresponse analysis is included in appendix A).

The questionnaire contained questions about health and medical consumption. In order to make the results of the survey comparable with other surveys, the Netherlands continuous Health Interview Survey was used as a guide for constructing the questionnaire (Mootz and Van den Berg, 1989). The questions with respect to health status referred to perceived health status, chronic conditions, functional disabilities and psychological unwell-being. The questionnaire also contained questions about date of birth, sex, the respondent's country of birth and that of his parents, education, marital status and the number of persons in the household. The questions about medical consumption referred to the consultation of a general practitioner, a medical specialist, physiotherapist, speech therapist and dentist, hospitalizations and the use of prescribed drugs. Table 3.3 gives an overview of the topics in the health survey.

The results for several questions about medical consumption could be compared with the results of the Netherlands continuous Health Interview Survey in 1992 for sickness fund members (Swinkels, 1994). This comparison was made for the consultation of a general practitioner, a specialist and alternative practitioners, visits to the dentist and the number of persons with complete dentures, use of physiotherapy and hospital admissions. Some small differences occurred. Compared to the national data relatively more members of Zorg and Zekerheid used physiotherapy

Table 3.3 Topics in the health survey

Topics	number of items
Health:	
* perceived health	1
* perceived health 5 years ago	1
* number of days in bed because of illness or injury during last 6 months	1
* chronic diseases during the last 5 years	24 (adults) 16 (children)
* height and weight, to construct a body-mass index	2
Adults only:	
* perceived health status (scale)	23
* functional disabilities in communication and mobility	8
* disabilities in activities for daily living	3
* psychological unwell-being (Affect Balance Scale, negative items)	5
Medical consumption:	
* consultation of a general practitioner during last two months	1
* consultation of a specialist during last year	1
* visiting a dentist during last year, in combination with using dentures	2 (adults) 1 (children)
* admission to the hospital during last year	1
* use of physiotherapy during last year	1
* use of speech therapy during last year	1
* consultation of alternative practitioners during last year	7
* use of ambulatory mental care (Riagg) during last 5 years	1
* use of other health services during last year	3
* use of prescribed drugs during last two weeks	1
* the kinds of prescribed drugs used	16
* use of drugs without a prescription during last 2 weeks	1
Adults only:	
* health services use propensity (scale)	5

Table 3.3 Continued

Topics	number of items
Expected health care use:	
* consultation of a specialist next year	1
* admission to the hospital next year	1
Demographic information:	
* sex	1
* date of birth	1
* country of birth of respondent	1
* country of birth of respondents parents	2
* education (adults), education of parent (children)	1
* marital status (adults), marital status of parents (children)	1
* number of persons in household	1
* number of children in the household	1

and had a hospital admission; relatively fewer Zorg and Zekerheid members consulted a specialist and used complete dentures.

The answers to the survey questions about contact with a specialist, a physiotherapist, a speech therapist and hospital admissions during the last year could also be compared with administrative data from 1992. Although there is a small time lag, the agreement between survey and administrative data was high for contact with a physiotherapist and a specialist, for both adults and children. For contact with a speech therapist and hospital admissions the agreement between the data sources was low. The low agreement for contact with a speech therapist could be explained by incomplete administrative data. For hospital admissions the survey data gave higher prevalence estimates than the administrative data. For children this could be explained by the fact that parents gave affirmative answers when the child went to the hospital for day case treatment. Confusion of day case treatment in the hospital with hospitalizations was for adults only partly an explanation for the differences in prevalences between the two data sources. Another explanation was recall bias resulting in an overestimation of admissions (see appendix B for an extensive comparison

of survey and administrative data). This overestimation of hospital admissions for the survey data explained the difference in admission rate between the survey among Zorg en Zekerheid members and the Netherlands continuous Health Interview Survey.

3.2 Method

This section describes the method for estimating and predicting health care expenditures, which is used throughout this thesis. In some of the next chapters specific and additional analyses will be presented. The methods used for these analyses are described in the chapters concerned.

3.2.1 *Models*

The first regression model to be estimated by ordinary least squares (OLS) is a basic capitation payment model, the so-called 'demographic' model. This model serves as a reference model and comprises only simple risk factors -i.e. age, sex, region, disability and employment- as independent variables and health care expenditures in year t as dependent variable. Age and sex are included in the demographic model in the form of 35 dummy variables (18 five-year age groups for each sex minus 1). The reason for distinguishing so many age/sex groups is that the present study is based on a sample of the *general* population where a linear effect of age on health care expenses is unrealistic. Moreover, national statistics show that in both the fertile years and in the age group beyond 65 years of age health care expenses for men and women are quite different. The address density of the surrounding area is used as region variable. This factor measures degree of urbanization and has five categories (Den Dulk et al., 1992). Disability is included in the demographic model as a dummy variable indicating whether or not the compulsory cause for enrolment with the sickness fund is 'being a recipient of disability benefits'; employment as a dummy

variable indicating whether or not the cause for enrolment is 'being a wage earner'.

Subsequently, DCG regression models that comprise the variables from the demographic model together with dummy variables for DCGs, are estimated by OLS. The dummies for the year t-1 DCGs are incorporated in the one-year DCG model. In the chapters 8 and 9 the one-year DCG model will be extended to a multi-year DCG model. For the two-year DCG model, the one-year DCG model is extended with dummies for the DCGs in the year t-2, whereas the three-year DCG model comprises dummies for the year t-3 DCGs as well. Table 3.4 gives an overview of independent variables in the demographic model and the DCG models.

Table 3.4 Description of alternative models to predict costs in year t

Model	Risk-adjusters
Demographic	35 (2x18-1) age/sex dummies + 1 dummy for disability + 1 dummy for employed + 4 dummies for region
One-year DCG	variables of the demographic model + dummies for DCGs in year t-1
Two-year DCG	variables of the demographic model + dummies for DCGs in year t-1 + dummies for DCGs in year t-2
Three-year DCG	variables of the demographic model + dummies for DCGs in year t-1 + dummies for DCGs in year t-2 + dummies for DCGs in year t-3

For the estimation of the various capitation models in the chapters 4 to 6 and chapter 8 health care expenditures in 1992 will be used as the dependent variable; in chapter 9 expenditures in 1994. In chapter 4 for

DCGs the original DCG classification as suggested by Ash and coworkers (1986, 1989) will be used. In latter chapters also dummies for DCGs based on other DCG classifications will be used in the regression models.

3.2.2 *Estimation*

The models are assumed to be linear in the coefficients and all include an intercept. They are estimated by means of ordinary least squares, with an individual's annual health care expenditures in year t as dependent variable and the various sets of risk-adjusters as independent variables. This statistical specification closely follows the cell-based approach which is essentially the form in which the actual capitation payments are calculated for the sickness funds in the Netherlands.

One of the problems with predicting health care expenditures stems from the non-normal shape of the distribution of annual expenditures. Most people have little or no health care expenditures in a particular year and a very small percentage have extremely high use (Berk and Monheit, 1992; Russell and Chaudhuri, 1992). Table 3.5 gives an illustration of the distribution of actual costs and the costs predicted by means of the demographic model. It shows for example that the 10% members with the highest cost in 1992 (above 1,888 guilders [12]) are responsible for 80% of all health care expenditures while they would contribute only 15% (demographic model) to the capitation payment of their sickness fund. Duan et al. (1983) developed several versions of multi-part models to capture skewness. Some of the models in the present study were also estimated by means of a two-part model. In the first part a probit analysis is used to estimate the probability that a person will have above zero health care expenditures in year t. In the second part the logarithm of positive cost is estimated by means of ordinary least squares. The predictive accuracy of this two-part model was comparable to that of the more simple linear model (Lamers and Van Vliet, 1996). This is in line with the findings of other researchers (Hornbrook and Goodman, 1995).

[12] One Dutch guilder is approximately 0.5 US$ in september 1997.

Since the linear model is directly transferable to the cell-based approach which is the form in which the actual capitation payments are calculated and since sickness funds are paid guilders and not log transformed guilders, the results presented are estimated with the linear model.

Table 3.5 Actual and predicted costs in 1992 for groups formed on costs in 1992

		percentage of costs in 1992	
percent of members ranked by 1992 costs	threshold	actual costs	predicted costs: demographic model
Top 1 percent	21,468	34.6 %	2.1 %
Top 5 percent	4,897	66.9 %	8.8 %
Top 10 percent	1,888	79.9 %	15.4 %
Top 25 percent	591	91.8 %	30.5 %
Top 50 percent	156	98.3 %	54.6 %

Since 1992 members may leave the cohort. For those who left the sickness fund during a year, the costs are raised to annual rates. At the same time weights are assigned for the part of the year they were in the data set. This means that for someone who died at the end of March 1992 and who had 3,000 guilders health care costs during the first three months of 1992, the annual rate will become 12,000 guilders and the assigned weight 3/12. By applying this procedure mean costs per person-year for the total data set are not changed [13].

To assess the accuracy of the capitation models for predicting future costs a split-sample method is applied, whereby the data set is divided randomly into two halves, labelled the 'estimation data set' and the 'prediction data set'. The models are fitted to the estimation data set and the estimated

[13] As a result of raising costs to annual rates the costs in 1992 of a few persons who died in the first months of the year became extremely high. For the estimation and prediction of costs in 1992 the (weighted) ten persons with the highest costs were excluded.

coefficients are then used to calculate predicted costs in the prediction data set. This cross-validation approach reduces the possibility of over-fitting, both deliberate – inclusion of ever more explanatory variables will inevitably increase R^2-values – and by chance – outliers as high as 200 times the average are rare but possible in health care expenditures data and may have a substantial impact on estimated models. The ability of alternative capitation models to predict future costs is evaluated in the prediction data set by means of R^2-values and the deviations of predicted from actual costs for various groups of members. The use of R^2 as a measure for the predictive accuracy of a capitation model is common. Most researchers provide R^2-values. As a measure of prediction bias, some researchers calculate predictive ratio's (for example, Ash et al., 1989, Ellis et al., 1996, Weiner et al., 1996 and Gruenberg et al., 1996). Predictive ratio's evaluate how closely each model predicts the costs for different subgroups. These predictive ratio's are calculated for groups based on demographic factors like age and sex, random and nonrandom subgroups. Most interesting are nonrandom groups of persons with predictable high expenditures such as groups biased on prior use. In this study actual costs will be compared with the predictions of various capitation models for groups based on prior utilization, i.e. DCGs and prior costs. The prior utilization groups are based on DCGs and costs in 1988 to avoid using the same DCGs as an independent variable in the regression models and forming subgroups with the same information. In case of such an overlap the predicted costs by a DCG model will not differ much from the actual costs for the subgroups based on the same DCGs. For example a DCG model using 1991 DCGs as a risk-adjuster to predict costs in 1992 will make predictions close to actual costs for subgroups based on 1991 DCGs. In chapter 8 a three-year DCG model including as a risk-adjuster diagnostic information from hospitalizations in the three-year period 1989-1991, will be used to predict health care expenditures in 1992. Therefore, the subgroups for which predicted costs are compared with actual costs have to be based on utilization in 1988 to avoid overlap.

4

The US Diagnostic Cost Group model applied to Dutch sickness fund data

4.1 Introduction

The DCG model was developed with and applied to data from the US Medicare program, which mainly covers people of 65 years and over. In the Netherlands, Van Vliet and Van de Ven (1993) applied the DCG model to a database of persons of all ages, who had private health insurance. For this population the predictive accuracy of the DCG was substantially higher than that of a demographic model. In this chapter the DCG model is applied to Dutch sickness fund data of both *aged and non-aged* people.

4.2 The original US Diagnostic Cost Group classification

Since people with serious chronic diseases have a relatively high probability of being hospitalized in a given period, hospitalization can be used as an indicator of poor health. Ash and several coworkers developed various variants of Diagnostic Costs Groups (DCGs), using both clinical and economic criteria. Ash et al. (1986, 1989) first classified diagnoses into 78 clinically homogeneous groups. These were further clustered into 9 groups according to the empirically determined similarities in the *future* costs of individuals hospitalized with different diagnoses. Subsequently, those groups were used in the form of 9 dummy variables -those without an admission comprising the reference group- in a regression model explaining annual health care costs of individuals. This DCG model gave an increase in R^2 of about 4 percentage points compared to a model that comprised

only age, sex and welfare status (0.5% versus 4.7%). The costs in the following year of those individuals who had had a hospital admission appeared to be three times as high as the costs of those without admission.

Sickness fund members in the panel data set were assigned to DCGs on the basis of the diagnoses from hospital admissions from 1988 to 1993. For each year separately the diagnoses were classified in one of the 9 original DCGs developed by Ash et al. (1986, 1989). Appendix C provides an overview of ICD-9-CM codes per DCG. People without hospital admission were assigned to group 0. Groups 8 and 9 were combined because group 9 contains few observations in our data set. For people with more than one hospital admission in a certain year the diagnosis classified in the highest DCG was used. Table 4.1 lists some common diagnoses within each

Table 4.1 Examples of ICD-9-CM diagnoses falling into each DCG (DCG classification according to Ash et al.)

DCG	ICD codes	Description	% of admissions in 1988
1	630 - 676	Pregnancy-related problems	
	540 - 542	Appendicitis	9.7 %
2	366	Cataract	
	470 - 478	Other diseases of respiratory tract	
	550 - 553	Hernia of abdominal cavity	9.2 %
3	800 - 839	Injuries involving fractures and dislocations	3.3 %
4	740 - 759	Congenital anomalies	0.9 %
5	410 - 414	Ischemic heart disease	
	415 - 417	Diseases of pulmonary circulation	
	785	Cardiovascular symptoms	6.7 %
6	250	Diabetes mellitus	1.1 %
7	490 - 496	Chronic obstructive pulmonary disease and like conditions	1.4 %
8-9	428	Heart failure	
	141 - 149 and 200 - 208	Various malignant neoplasms	1.2 %

DCG. About one third of the hospital admissions is related to the diagnoses listed in table 4.1.

Table 4.2 gives for 1988 the distribution of the individuals in the data set over the 9 DCGs, as well as their mean costs in 1989 and 1992. The distribution in table 4.2 closely resembles the distribution of diagnoses of all hospital admissions in the Netherlands (SIG, 1990), supporting the generalizibility of the findings. Mean costs are expressed as ratios with regard to the overall mean. The costs of the 93.3% of the persons without admissions in 1988 are on average 20% below the overall mean in the next year (95% confidence interval ranging from 0.76 to 0.84) and still 9% below the overall mean after four years (95% confidence interval ranging from 0.87 to 0.95).

Table 4.2 For groups formed on DCGs in 1988: mean costs in 1989 and 1992 expressed as a ratio with regard to overall mean costs

		mean costs in 1989		mean costs in 1992	
DCG in 1988	% [a]	ratio	95% - confidence interval	ratio	95% - confidence interval
1	15.4	1.53	1.29 - 1.77	1.14	0.98 - 1.30
2	12.3	1.65	1.41 - 1.89	1.54	1.32 - 1.76
3	20.4	3.16	2.71 - 3.61	1.72	1.52 - 1.92
4	16.6	4.10	3.59 - 4.61	2.59	2.30 - 2.88
5	28.9	5.34	4.91 - 5.77	2.85	2.63 - 3.07
6	1.9	8.55	6.55 - 10.55	4.73	3.65 - 5.81
7	3.2	7.27	5.92 - 8.62	4.70	3.68 - 5.72
8-9	1.3	11.18	8.63 - 13.73	7.90	5.96 - 9.84

[a] Percentage of people hospitalized in 1988.

Table 4.2 shows that people with an admission in one year and thus assigned to one of the DCGs, have costs in the next year that are at least 53% above the overall mean. For DCG 3 the costs are about three times the mean, while for the higher DCGs this figure can become seven and eleven. Four years after an admission that falls in DCG 3 costs are still

70% above the overall mean, even more for the higher DCGs. The next section describes the predictive accuracy of the one-year DCG model with the original DCG classification of Ash and coworkers.

4.3 Predictive accuracy

To assess the predictive accuracy of the demographic model and the one-year DCG model ordinary least squares regression analyses were performed with an individual's annual health care expenditures in 1992 as the dependent variable and demographic factors and DCGs in 1991 as the independent variables (see table 3.4). Table 4.3 presents the R^2-values for the estimated models in the prediction data set for the non-aged and the elderly separately as well as together. The results show that the predictive accuracy of the one-year DCG model, in terms of R^2, is more than twice that of the demographic model. For the elderly the R^2-values for the demographic and DCG model are consistent with findings of Ash et al. (1986, 1989).

Table 4.3 R^2-values * 100 for prediction for alternative capitation models per age group

	young (< 65 years)	elderly (≥ 65 years)	total
Demographic	1.66	0.33	3.07
One-year DCG	4.58	4.52	6.35

Table 4.4 shows the estimated coefficients of the DCGs for non-aged and elderly separately as well as combined. Admissions in 1991 belonging in DCG 1 do not contribute significantly in explaining costs in 1992. For most DCGs the coefficients for the elderly are higher than for the group of non-aged persons. One can expect a more or less increasing trend in

coefficients: the higher the DCG, the higher the coefficient. The coefficients of the DCGs for the two age groups combined resemble this trend best. When capitation payments are based on a one-year DCG, a coefficient for DCG 4 of 4272 (last column) means that next year's capitation payment to a sickness fund for a person with a admission to the hospital for a diagnosis belonging to DCG 4 is 4272 guilders higher than for a person with the same demographic characteristics without an admission.

Table 4.4 Coefficients of DCGs from estimating costs in 1992 by the one-year DCG model per age group

	Coefficients		
DCGs in 1991	young (< 65 years)	elderly (≥ 65 years)	total
DCG 1	228 [ns]	- 1423 [ns]	196 [ns]
DCG 2	626 [ns]	2020	1066
DCG 3	1205	5241	2112
DCG 4	4799	3448	4272
DCG 5	4679	7880	5612
DCG 6	3155	6833	5020
DCG 7	3445	8796	5811
DCG 8 - 9	11877	11034	11394
diagnosis unknown	1502	4050	2347

[ns] The coefficient is not statistically significant from 0 ($p > 0.05$).

The predictive accuracy of the regression models is also evaluated more directly, namely by assessing how well they predict costs for various relevant subgroups in the data set. So-called cost ratios for the year 1992 were determined for various subgroups defined on components of health care costs and utilization in 1988. The cost ratio for a particular group and a given model is formed by dividing the costs actually incurred for the individuals in that group by the costs predicted by the model. The relevance of an analysis of cost ratios is that sickness funds can try to

'game' the capitation model by using information not (completely) included in the model as a vehicle for risk selection. Cost ratios indicate the average profit a sickness fund may expect to make by attracting people from groups in which average costs are below their capitation payment (ratio < 1) and by rejecting people for whom the opposite holds (ratio > 1). Table 4.5 presents the cost ratios for the elderly; table 4.6 for persons younger than 65 years of age and table 4.7 for the two age groups combined.

Table 4.5 Ratios of actual to predicted costs in 1992 for various subgroups formed on prior use and costs for the *elderly* (persons ≥ 65 years)

	1988		Ratio = actual / predicted costs in 1992		
Risk factor / interval	N (%)	mean costs in Dfl.	no risk-adjustment[a]	demo-graphic	one-year DCG
DCG in 1988					
0 (no admission)	88.5	307	0.89	0.90	0.91[ns]
1 - 2	2.1	6635	1.30	1.22[ns]	1.16[ns]
3 - 4	3.9	10985	1.69	1.63	1.45
5	4.3	14243	1.92	1.87	1.59
6 - 9	1.2	16723	2.85	2.68	2.14
Total costs in 1988					
0 Dfl.	38.6	0	0.57	0.57	0.61
1 - 3200 Dfl.	50.2	449	1.13	1.14[ns]	1.12[ns]
> 3200 Dfl.	11.2	12804	1.91	1.83	1.60
Outpatient costs in 1988					
0 Dfl.	39.0	85	0.57	0.57	0.61
1 - 1000 Dfl.	48.8	1081	1.12	1.13	1.11[ns]
> 1000 Dfl.	12.2	9056	1.91	1.82	1.62

[a] In this case predicted costs is the population average in 1992: 3215 Dutch guilders.
[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

Table 4.6 Ratios of actual to predicted costs in 1992 for various subgroups formed on prior use and costs for persons *young*er than 65 years

	1988		Ratio = actual / predicted costs in 1992		
Risk factor / interval	N (%)	mean costs in Dfl.	no risk-adjustment[a]	demographic	one-year DCG
DCG in 1988					
0 (no admission)	93.9	209	0.93	0.94[ns]	0.96[ns]
1 - 2	1.8	4637	1.33	1.22	1.16[ns]
3 - 4	2.3	6513	1.62	1.43	1.26
5	1.7	11630	2.65	2.03	1.65
6 - 9	0.3	12534	5.97	4.06	2.95
Total costs in 1988					
0 Dfl.	57.1	0	0.72	0.78	0.82
1 - 3200 Dfl.	38.3	471	1.21	1.13	1.10[ns]
> 3200 Dfl.	4.6	10451	2.63	2.03	1.72
Outpatient costs in 1988					
0 Dfl.	57.7	69	0.73	0.79	0.82
1 - 1000 Dfl.	36.8	768	1.20	1.13	1.10[ns]
> 1000 Dfl.	5.5	6289	2.49	1.84	1.62

[a] In this case predicted costs is the population average in 1992: 957 Dutch guilders.
[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

Among the elderly relatively more persons had a hospital admission in 1988 and relatively fewer persons had zero costs in 1988. As the low R^2-value for elderly suggested, the predictions of the demographic model give minor improvements in the costs ratios compared to the situation of 'no risk-adjustment' i.e. when the population average is used as the predicted costs. For elderly persons with a hospital admission in DCG 6-9 in 1988

the average costs in 1992 are 185% above the (elderly) population mean and 168% above the costs predicted with the demographic model [14].

Table 4.7 Ratios of actual to predicted costs in 1992 for various subgroups formed on prior use and costs for *both aged* and *non-aged* persons

	1988		Ratio = actual / predicted costs in 1992		
Risk factor / interval	N (%)	mean costs in Dfl.	no risk-adjustment [a]	demographic	one-year DCG
DCG in 1988					
0 (no admission)	93.2	221	0.91	0.93	0.94[ns]
1 - 2	1.9	4926	1.35	1.21	1.16[ns]
3 - 4	2.5	7432	1.87	1.49	1.32
5	2.0	12355	2.83	1.89	1.58
6 - 9	0.4	14117	5.60	3.09	2.37
Total costs in 1988					
0 Dfl.	54.7	0	0.64	0.73	0.77
1 - 3200 Dfl.	39.8	467	1.25	1.12	1.10
> 3200 Dfl.	5.5	11079	2.78	1.88	1.62
Outpatient costs in 1988					
0 Dfl.	55.3	71	0.64	0.74	0.78
1 - 1000 Dfl.	38.4	820	1.24	1.12	1.09
> 1000 Dfl.	6.3	6985	2.65	1.79	1.60

[a] In this case predicted costs is the population average in 1992: 1253 Dutch guilders.
[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

[14] Using the population average costs and the cost ratio for the situation of no risk adjustment, the average profit or loss in Dutch guilders per subgroup can be calculated. For example, the actual costs in 1992 for the subgroup of the elderly with a hospital admission in DCG 6-9 in 1988 (table 4.5) are 9163 Dutch guilders (= 2.85 (ratio for no risk adjustment) * 3215 (population average)). The costs predicted by the demographic model are 3419 guilders (9163/2.68 (cost ratio for demographic model)). So, in case of a demographic capitation model the average loss for the subgroup is 5744 guilders (9163 - 3419).

For the non-aged and the elderly separately as well as together the tables 4.5 to 4.7 reaffirm that the demographic model has poor predictive ability and may lead to risk selection. For people in DCG 6-9 (in 1988) the average costs in 1992 are 209% above the costs predicted by the demographic model. For people in DCG 0 (no hospital admission) the actual costs are 7% below the costs predicted by the demographic model. The difference between actual and predicted costs diminishes for the DCG-subgroups when costs are predicted by the DCG model, for example the above-mentioned difference between actual and predicted costs of 209% (demographic model) reduces to 137% when costs are predicted by the DCG model. For the non-aged these figures are 306% respectively 195%; for the elderly 168% respectively 114%. For the sickness fund members with the highest costs in 1988 the expected loss in 1992 reduces by 30% (non-aged), 28% (elderly) and 30% (total) when predicted costs are based on the one-year DCG model compared with the demographic model. For the group of persons with the highest outpatient costs in 1988 the reduction in expected loss is 26% (non-aged), 24% (elderly) and 24% (total) when predicted costs are based on the DCG model compared with the demographic model.

4.4 Conclusion

For the non-aged and the elderly separately as well as together the predictive accuracy of the risk-adjusted capitation model improved when the demographic variables were supplemented with diagnostic information from prior hospitalizations in the form of DCGs. The R^2-value of the one-year DCG model was more than twice the R^2-value of the demographic model. The expected loss in 1992 for members with a hospital admission in the highest DCGs in 1988 (total population) decreased from 209% (demographic model) to 137% (one-year DCG model).

These results show that the DCG model, developed with US data of people of 65 years and over can be applied to Dutch sickness fund data of

persons of all ages. In the next chapters DCG models will only be applied to data of non-aged and elderly persons together.

5

The development of a Diagnostic Cost Groups classification for the Dutch situation

5.1 Introduction

Ash and several coworkers (1986, 1989) developed the Diagnostic Costs Groups model with data from the US Medicare program, which mainly covers people of 65 years and over. First, they classified diagnoses into 78 clinically homogeneous groups, the so-called Diagnostic SubGroups (DSGs). These were further clustered into 9 groups according to the empirically determined similarities in the *future* costs of individuals hospitalized with different diagnoses. Future costs here are the costs in the year following a hospital admission.

In this chapter the development of a DCG classification based on Dutch cost data is described. Cluster analysis is performed to group the 78 DSGs into nine DCGs on the basis of various definitions of follow-up costs. In section 5.2 the data used in the cluster analysis and the method of clustering are described. In the next section the predictive accuracy of DCG models with different DCG classifications is compared. The DCG classifications used are the original classification of Ash et al. and three classifications developed with Dutch sickness fund data. One classification is based on follow-up costs in the first year after a hospital admission; another classification on age and sex corrected follow-up costs in the first year after admission and the third classification is based on costs in the second and third year after hospital admission. Section 5.4 investigates whether the nine DCGs can be further clustered into a smaller number without affecting the predictive accuracy. In section 5.5 the effect on the predictive accuracy of the DCG model is examined when the so-called 'chronicity of the diagnoses' is incorporated into the DCG model. The

chronicity of the diagnoses developed by Anderson, Steinberg, Powe et al. (1990) for the Payment Amount for Capitated Systems is used for this purpose. In the last section the conclusions about the development of a DCG classification for the Dutch situation are summarized.

5.2 Clustering Diagnostic Subgroups to Diagnostic Cost Groups

5.2.1 *Data*

The cluster analysis is applied to the panel data over the period 1988-1991. For this cluster analysis a data set is created with all admissions in 1988 as records. A person with more than one hospital admission in 1988 forms more than one record in this data set. When a person had more than one hospital admission in 1988 with diagnoses belonging to the same DSG, only the information of the last admission is used in the data set. Table 5.1 presents the number of hospital admissions per person in 1988 in the data set for the cluster analysis. For the cluster analysis the data are unweighted.

Table 5.1 Percentage of persons by number of hospital admissions in 1988 for diagnoses in different DSGs

	number of persons	percentage
1 hospital admission in 1988	15,468	87.8 %
2 hospital admissions in 1988	1,737	9.8 %
3 hospital admissions in 1988	328	1.9 %
4 hospital admissions in 1988	65	0.4 %
5 hospital admissions in 1988	18	0.1 %
Total	17616	100 %

The data set for the cluster analysis comprises information about 20,276 hospital admissions of 17,616 individuals. For each admission the following variables are available: the diagnosis in the form of an ICD-9-CM code, the DSG to which the diagnosis belongs, and the chronicity of the diagnosis, according to the chronicity classification of Anderson, Steinberg, Powe et al. (1990). Anderson and coworkers assigned a disease chronicity classification to each clinical disorder by asking 169 physicians in 31 specialties to rate each 5-digit ICD-9-CM diagnosis code in their specialty as being a chronic disorder, an acute disorder with possible sequelae or an acute self-limiting disorder.

Follow-up costs per year for 1989, 1990 and 1991 are matched with the data set of admissions in 1988. When someone had more than one hospital admission in 1988 and occurs more than once in the data set, the follow-up costs are divided by the number of admissions in 1988, resulting in *mean follow-up costs per admission.* Age and sex corrected follow-up costs for 1989 were also matched with the data set.

5.2.2 *Method*

The 20,276 admissions in the data set were first classified according to DSG. The 78 DSGs are the starting point for the cluster analysis. Three DSGs contain less than 20 admissions and were excluded from the cluster analysis. These three DSGs are classified into a DCG according to the classification of Ash et al.. The remaining 75 DSGs are clustered on the basis of follow-up costs into nine DCGs. For comparability with the original DCG classification of Ash et al. the cluster analysis stopped at nine DCGs. Appendix D provides an overview of the number of admissions and mean follow-up costs per DSG.

Per DSG corresponding to hospital admissions in 1988 the mean follow-up costs and the variance within the DSG are determined. The DSGs are clustered on mean follow-up costs. In the cluster analysis a correction is made for the number of admissions per DSG. Ward's hierarchical clustering method is used for the analysis. This method unions two clusters

at each step whose fusion results in the minimum increase in 'information loss', in terms of an error sum of squares criterion (Everitt, 1993).

5.3 DCG classifications based on different types of follow-up costs

Three cluster analyses are performed with three different types of follow-up costs. These follow-up costs are total health care costs of the individual concerned in the first year after his/her hospital admission, age and sex corrected costs in the first year after admission and costs in the second and third year after admission. Table 5.2 gives an overview of the follow-up costs used in the cluster analyses.

Table 5.2 Description of alternative DCG classifications based on different types of follow-up costs

DCG-Ash	original US classification, according to Ash et al.
DCG-Follow-Up Year one (*DCG-fuy 1*)	DCG classification based on follow-up costs in the first year after admission (1989)
DCG-Follow-Up Year one age/sex corrected (*DCG-fuy 1 age/sex corrected*)	DCG classification based on age and sex corrected follow-up costs in the first year after admission (1989)
DCG-Follow-Up Year two + three (*DCG-fuy 2+3*)	DCG classification based on follow-up costs in the second and third year after admission (1990 + 1991)

Since some health problems are age or to gender related, one DCG classification is based on age and sex corrected follow-up costs. The costs in the second and third year after hospital admission are used with the idea of

excluding follow-up costs that are related to self-limiting acute conditions. The chronic component of a health problem resulting in a admission to the hospital, will probably still generate health care expenditures after two or three years.

Table 5.3 shows the number of admissions per DCG for the various DCG classifications. Clustering on the basis of age and sex corrected costs in the first year after admission results in the largest number of admissions in DCG 1. This clustering into DCGs has a relatively small number of admissions in the higher DCGs. The large number of admissions in DCG 5 for the original DCG classification and the classification based on costs in the second and third year after admission is striking. For all the DCG classifications the 20 admissions of DSG 78 in the panel data set (Diseases of the Genitourinary System: Nephritis, nephrotic syndrome and nephrosis) form DCG 9.

Table 5.3 Number of admissions in 1988 per DCG for various DCG classifications

	DCG-Ash	DCG-fuy 1 *	DCG-fuy 1 age/sex corrected	DCG-fuy 2+3
DCG 1	2,489	3,439	6,022	1,939
DCG 2	2,597	2,122	2,606	2,775
DCG 3	4,142	3,610	3,698	3,161
DCG 4	3,617	3,155	2,409	2,076
DCG 5	6,073	2,856	3,296	6,517
DCG 6	402	3,426	1,255	2,313
DCG 7	632	678	895	1,032
DCG 8	299	965	70	438
DCG 9	20	20	20	20

* A description of the DCG classifications is given in table 5.2.

Table 5.4 shows on the DSG level and table 5.5 on the level of admissions the relation between the various DCG classification using Spearman rank correlation coefficients. The rank correlation coefficients on the admission level are higher than the coefficients on DSG level. The classification based on follow-up costs in the first year after admission and the age and sex corrected costs in that year are strongly related. The weakest relation occurs between the original DCG classification of Ash et al. and the classification based on costs in the second and third year after admission.

Table 5.4 Spearman rank correlation coefficients between various classifications of DSGs into nine DCGs (N=78 DSGs)

	DCG Ash	DCG-fuy 1	DCG-fuy 1 age/sex corrected
DCG-fuy 1 *	0.63		
DCG-fuy 1 age/sex corrected	0.68	0.96	
DCG-fuy 2+3	0.58	0.80	0.73

* A description of the DCG classifications is given in table 5.2.

Table 5.5 Spearman rank correlation coefficients between various classifications of admissions into nine DCGs (N=20.276 admissions in 1988)

	DCG Ash	DCG-fuy 1	DCG-fuy 1 age/sex corrected
DCG-fuy 1 *	0.76		
DCG-fuy 1 age/sex corrected	0.77	0.96	
DCG-fuy 2+3	0.68	0.84	0.77

* A description of the DCG classifications is given in table 5.2.

The three DCG classifications based on various follow-up costs are used in a one-year DCG model (see table 3.4). 1991 DCGs are used in the one-year DCG model. The predictive accuracy of these models, in terms of R^2-values, is shown in table 5.6. As a reference model the DCG model with the original DCG classification of Ash et al. is added to table 5.6. The differences in R^2-values between the models are small. The R^2-values for the models with the DCG classification based on Dutch follow-up costs are a little higher than for the model with the original DCG classification of Ash et al.. The model with the DCG classification based on follow-up costs in the second and third year after admission provides the highest R^2-value.

Table 5.6 R^2 x 100 for estimation and prediction of costs in 1992

	estimation	prediction
Demographic	4.34	3.07
One-year DCG:		
DCG Ash	7.71	6.35
DCG-fuy 1 *	8.15	6.49
DCG-fuy 1 age / sex corrected	8.47	6.34
DCG-fuy 2+3	8.66	6.84

* A description of the DCG classifications is given in table 5.2.

The predictive accuracy of the DCG models with the different DCG classifications is also evaluated by determining cost ratios for various subgroups defined on components of health care costs and utilization in 1988. Table 5.7 presents these costs ratios.

The differences between the cost ratios for the one-year DCG models with various DCG classifications are small. The largest differences occur for the subgroup with a hospital admission in the highest DCGs in 1988. For this subgroup the models with the DCG classifications based on age and sex

corrected follow-up costs and on follow-up in the second and third year after admission show cost ratios closest to 1. For the subgroups defined on total and outpatient costs in 1988 the ratios of the models with DCG classifications based on different types of follow-up costs are almost identical.

Table 5.7 Ratios of actual to predicted costs in 1992 for various subgroups formed on prior use and costs for one-year DCG models with various DCG classifications

	Ratio = actual / predicted costs in 1992			
Risk group /interval	DCG Ash	DCG-fuy 1 *	DCG-fuy 1 age/sex corrected *	DCG-fuy 2+3 *
DCG (Ash) in 1988				
0 (no admission)	0.94^{ns}	0.95^{ns}	0.95^{ns}	0.95^{ns}
1-2	1.16^{ns}	1.16^{ns}	1.15^{ns}	1.14^{ns}
3-4	1.32	1.32	1.31	1.31
5	1.58	1.55	1.54	1.56
6-9	2.37	2.27	2.23	2.23
Total costs in 1988				
0 Dfl.	0.77	0.77	0.77	0.77
1-3200 Dfl.	1.10	1.11	1.11	1.11
> 3200 Dfl.	1.62	1.60	1.59	1.59
Outpatient costs in 1988				
0 Dfl.	0.78	0.77	0.77	0.77
1-1000 Dfl.	1.09	1.10	1.10	1.10
> 1000 Dfl.	1.60	1.57	1.56	1.57

* A description of the DCG classifications is given in table 5.2.

ns No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

Summarizing, the differences in predictive accuracy, in terms of R^2-values and cost ratios, between DCG models using different DCG classifications

are small. The model with the DCG classification based on follow-up costs in the second and third year after hospital admission yields the highest R^2-values. As far as cost ratios are concerned, the predictive accuracy of the models with DCG classifications based on different types of follow-up costs are almost identical. From a theoretical point of view a DCG classification based on follow-up costs in the second and third year after admission is preferred. By not using the costs in the first year after hospital admission, follow-up costs that are related to self-limiting acute conditions are excluded. The resulting DCG classification is based on more chronic health problems, which after two or three years still generate health care expenditures. Therefore, in the next section this DCG classification will be used.

5.4 A further reduction of the nine DCGs

This section analyses whether the nine DCGs can be clustered further into a smaller number without affecting the predictive accuracy. The DCG classification used in this section is based on follow-up costs in the second and third year after hospital admission.

To compare the results of the cluster analyses with a decreasing number of clusters (DCGs) with each other, ordinary least squares regressions are performed. For these analyses the data set with all hospital admissions in 1988 as records is used (see section 5.2.1). In the regression model the follow-up costs in 1990 plus 1991 are used as the dependent variable and various number of dummy variables for DCGs as the independent variables. As a reference model a regression analysis is performed with a dummy variable for each of the 75 DSGs as independent variables, the situation without clustering DSGs into a smaller number of DCGs. Next, a F-test on linear restrictions is performed. This test examines whether a reduction of the 75 DSGs into a restricted number of DCGs results in a loss of predictive power. Table 5.8 shows the test results. These test results show that the number of DCGs can be reduced from nine to five DCGs.

This reduction in the number of DCGs should not affect the predictive accuracy of the DCG model. In the DCG classification with five DCGs, the DCG 1, 2, 3 and 4 from the classification in nine DCGs are clustered into a single DCG. This also holds for DCG 6 and DCG 7 from the DCG classification with nine DCGs (see table 5.3, last column).

Table 5.8 Test results (F-test) loss of predictive power for reduction of the number of DCGs

	F-statistic	degrees of freedom	p-value
reduction to:			
9 DCGs	0.132	66, 20179	$p > 0.25$
8 DCGs	0.188	67, 20179	$p > 0.25$
7 DCGs	0.346	68, 20179	$p > 0.25$
6 DCGs	0.569	69, 20179	$p > 0.25$
5 DCGs	0.885	70, 20179	$p > 0.25$
4 DCGs	1.579	71, 20179	$p < 0.01$
3 DCGs	2.815	72, 20179	$p < 0.01$

The R^2-values for DCG models with DCG classification with five and nine DCGs are shown in table 5.9. The R^2-values of the model with the DCG classification in five DCGs are somewhat smaller than for the model with a DCG classification in nine DCGs [15].

[15] The one-year DCG model with the DCG classification according to Ash et al. on the one hand and the DCG model with the DCG classifications in five DCGs based on follow-up costs in the second and third year after admission on the other hand can be seen as alternative nonnested hypotheses. These hypotheses were tested (Judge et al., 1980). The outcome of this test was that both hypotheses were rejected. So, this test does not give an empirical answer to the question which of the two models is better in estimating next year's health care expenditures.

Table 5.9 R^2 x 100 for estimation and prediction of costs in 1992

	estimation	prediction
One-year DCG		
DCG classification		
with 9 DCGs	8.66	6.84
with 5 DCGs	8.38	6.60

Table 5.10 shows the estimated coefficients of the nine and five DCGs by the one-year DCG models. Admissions in 1991 belonging to DCG 1 and 2 from the classification in nine DCGs do not contribute significantly in estimating costs in 1992. The DCGs from the DCG-classification in five DCG resembles best the increasing trend in coefficients: the higher the DCG, the higher the coefficient.

Table 5.10 Coefficients of DCGs from estimating costs in 1992 by one-year DCG models with DCG classifications with various numbers of DCGs

DCG-fuy 2+3 (1991)	Coefficients	DCG-fuy 2+3 (1991)	Coefficients
DCG 1	99 [ns]	DCG 1	957
DCG 2	463 [ns]		
DCG 3	1739		
DCG 4	1660		
DCG 5	3426	DCG 2	3420
DCG 6	7373	DCG 3	8949
DCG 7	13284		
DCG 8 - 9	9692	DCG 4 - 5	9680
diagnosis unknown	2365	diagnosis unknown	2361

[ns] The coefficient is not statistically significant from 0 ($p > 0.05$).

Table 5.11 presents cost ratios for various subgroups defined on components of health care costs and utilization in 1988 for the DCG model with a DCG classification in nine DCGs and one in five DCGs. The cost ratios for both models are almost identical. The predictive accuracy, in terms of cost ratios, is not affected by reducing the number of DCGs from nine to five. In the next section this DCG classification in five groups will be used together with the original DCG classification of Ash et al..

Table 5.11 Ratios of actual to predicted costs in 1992 for various subgroups formed on prior use and costs for one-year DCG models with DCG classifications with various numbers of DCGs

	Ratio = actual / predicted costs in 1992	
Risk group /interval	DCG-fuy 2+3 in 9 DCGs	DCG-fuy 2+3 in 5 DCGs
DCG (Ash) in 1988		
0 (no admission)	0.95[ns]	0.95[ns]
1-2	1.14[ns]	1.13[ns]
3-4	1.31	1.32
5	1.56	1.56
6-9	2.23	2.25
Total costs in 1988		
0 Dfl.	0.77	0.77
1-3200 Dfl.	1.11	1.11
> 3200 Dfl.	1.59	1.59
Outpatient costs in 1988		
0 Dfl.	0.77	0.77
1-1000 Dfl.	1.10	1.10
> 1000 Dfl.	1.57	1.58

[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

5.5 DCG models using chronicity of diagnoses

This section examines whether or not the predictive accuracy of DCG models can be improved by using information on the chronicity of the diagnoses. As a measure of chronicity of a diagnosis the chronicity classification as developed by Anderson, Steinberg, Powe et al. (1990) is used. Based on clinical judgement of 169 physicians in 31 specialties, they classified each diagnosis as belonging to a chronic disorder, an acute disorder with possible sequelae or an acute self-limiting disorder. Table 5.12 shows for the hospital admissions in 1988 in the data set, the chronicity of the diagnoses and whether or not the admission was for surgery. Almost half of the number of admissions were for chronic disorders; 14% of the admissions concerned acute disorders with possible sequelae. Less than 40% of the admissions were for surgery. In 1993 in the Netherlands 44% of all hospital admissions were for surgery (SIG, 1994). Most surgery occurs during hospital admissions for chronic disorders.

Table 5.12 Hospital admissions in 1988 by chronicity and surgery

	without surgery	with surgery	total N	total %
Acute	5,717	2,188	7,905	39,0 %
Sequelae	2,068	783	2,851	14,1 %
Chronic	4,644	4,873	9,517	46,9 %
Total	12,429	7,844	20,273	100 %

missing=3

The information about the chronicity of the diagnoses can be used in the DCG model in two manners. A dummy variable for whether or not the diagnosis is for a chronic disorder can be added to the one-year DCG model. This dummy variable is an additional risk-adjuster besides the dummy's for DCGs in the capitation model. The chronicity classification

can also be used in the clustering of DSGs to DCGs. For this cluster analysis each DSG is split up into diagnoses belonging to acute disorders (acute disorders with possible sequelae included) and diagnoses belonging to chronic disorders. The DSGs are split up when both categories, acute and chronic, have at least 20 admissions. For 37 DSGs this condition is met. In the cluster analysis 112 (75 plus 37) split DSGs are clustered into five DCGs on the basis of follow-up costs in the second and third year after hospital admission (costs in 1990 plus 1991).

Table 5.13 presents the number of admissions in 1988 per DCG for the DCG classification regardless of chronicity and the DCG classification based on chronicity. The number of admissions in DCG 2 is much larger for the DCG classification based on chronicity. The DCG classification, regardless of chronicity, has relatively many admissions in DCG 3.

Table 5.13 Number of hospital admissions in 1988 for a DCG classification using chronicity and a DCG classification regardless of chronicity

	DCG classification regardless of chronicity	DCG classification based on chronicity
DCG 1	9,951	7,357
DCG 2	6,517	11,340
DCG 3	3,345	1,018
DCG 4	438	536
DCG 5	20	20

R^2-values for one-year DCG models using information about the chronicity of the diagnoses and for a reference model, which did not use this information, are presented in table 5.14. The R^2-values are determined for DCG models with the original DCG classification of Ash et al. and classification in five DCGs based on follow-up costs in the second and third year after admission. Using chronicity as a dummy variable in the DCG model hardly affects the R^2-values. The coefficient for this dummy variable for the DCG model with the DCG classification based on follow-up costs in the second and third year after admission did not significantly

differ from zero. Using the DCG classification, which used chronicity in the clustering of (split) DSGs into DCGs, gives no improvements in R^2-value.

Table 5.14 R^2 x 100 for estimation and prediction of costs in 1992

	DCG Ash (nine DCGs)		DCG-fuy 2+3 (five DCGs)	
	estimation	prediction	estimation	prediction
One-year DCG				
DCG classification regardless of chronicity	7.71	6.35	8.38	6.60
Chronicity as dummy variable	7.73	6.43	8.39	6.65
DCG classification based on chronicity	not applicable	not applicable	8.42	6.56

Again, cost ratios are determined for the various subgroups defined on components of health care costs and utilization in 1988. Table 5.15 shows the cost ratios for the DCG models which used information about chronicity of the diagnoses and the reference model, that did not use this information. As the R^2-values already suggest, adding a dummy variable for chronicity to the DCG model, results in the same cost ratios as those of the DCG model without this dummy variable. This holds for both the DCG model with the original classification of Ash et al. and the model using DCGs based on follow-up costs in the second and third year after admission. Only the cost ratio for the subgroup with an admission in the highest DCGs in 1988 dropped a little for the DCG model with the DCG classification using chronicity in the clustering of (split) DSGs into DCGs.

Summarizing, the predictive accuracy of DCG models does not improve when the DCG model uses information on the chronicity of the diagnoses as defined in the chronicity classification of Anderson, Steinberg, Powe et al. (1990).

Table 5.15 Ratios of actual to predicted costs in 1992 for various subgroups formed on prior use and costs for one-year DCG models with DCG classifications with and without chronicity

	Ratio = actual / predicted costs in 1992				
	DCG classification: Ash (nine DCGs)		DCG Classification: follow-up year 2+3 (five DCGs)		
Risk group /interval	regardless of chronicity	chronicity as a dummy	regardless of chronicity	chronicity as a dummy	based on chronicity
DCG (Ash) in 1988					
0 (no admission)	0.94[ns]	0.94[ns]	0.95[ns]	0.95[ns]	0.95[ns]
1 - 2	1.16[ns]	1.16[ns]	1.13[ns]	1.14[ns]	1.14[ns]
3 - 4	1.32	1.32	1.32	1.32	1.32
5	1.58	1.58	1.56	1.55	1.57
6 - 9	2.37	2.37	2.25	2.25	2.16
Total costs in 1988					
0 Dfl.	0.77	0.77	0.77	0.77	0.77
1 - 3200 Dfl.	1.10	1.10	1.11	1.11	1.11
> 3200 Dfl.	1.62	1.62	1.59	1.59	1.60
Outpatient costs in 1988					
0 Dfl.	0.78	0.78	0.77	0.77	0.77
1 - 1000 Dfl.	1.09	1.09	1.10	1.10	1.10
> 1000 Dfl.	1.60	1.60	1.58	1.58	1.58

[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

5.6 Conclusion

In this chapter the development of a DCG classification based on Dutch cost data was described. The predictive accuracy of DCG models with the original DCG classifications of Ash et al. and classifications based on

different types of follow-up costs were compared. The differences in predictive accuracy, in terms of R^2-values and cost ratios, between DCG models using these alternative DCG classifications were small. The model with the DCG classification based on follow-up costs in the second and third year after hospital admission yielded the highest R^2-values. As far as cost ratios were concerned, the predictive accuracy of the models with DCG classifications based on alternative types of follow-up costs were almost identical. From a theoretical point of view a DCG classification based on follow-up costs in the second and third year after admission was preferred. By not using the costs in the first year after hospital admission, follow-up costs that are related to self-limiting acute conditions were excluded. The resulting DCG classification was based on health problems that are more likely to be chronic.

Initially the DCG classifications based on follow-up costs in the second and third year after admission had nine DCGs. These nine DCGs could be further clustered into five DCGs without affecting the predictive accuracy of the DCG model.

The effect on the predictive accuracy of using the chronicity of the diagnoses in the DCG model was examined. As a measure of chronicity of a diagnosis the chronicity classification as developed by Anderson, Steinberg, Powe et al. (1990) was used. This measure of chronicity did not contribute significantly to explaining next year's health care expenditures. This raises questions about the validity of this chronicity measure when applied to Dutch sickness fund data of persons of all ages. The predictive accuracy of DCG models could not be improved when the DCG model used information on the chronicity of the diagnoses.

The DCG classification for the Dutch situation developed with Dutch sickness fund data of both aged and non-aged persons is a classification with five DCGs based on follow-up costs in the second and third year after hospital admission (see appendix C for an overview of DSGs and diagnoses per DCG). This DCG classification does not explicitly take into account the chronicity of the diagnoses. The DCG model with the DCG classification developed for the Dutch situation has -when used on the

Dutch sickness fund database- a slightly higher predictive accuracy than the model with the original US DCG classification. Table 5.16 summarizes the differences in data used and clustering into DCGs between the original DCG classification developed by Ash et al. and the classification developed in the present study.

Table 5.16 Differences between the study of Ash et al. and the present study with regard to data and the clustering of diagnoses into DCGs

	study of Ash et al.	present study
Data:	- from *US* Medicare program - of persons *65 years and over* - 1979-1980 data	- *Dutch* sickness fund data - of persons of *all ages* - 1988-1991 data
Clustering:	- on follow-up costs *next year* - on *total* follow-up costs - into *9* DCGs	- on follow-up costs in *second and third year* after hospital admission - on *mean* follow-up costs per admission - into *5* DCGs

The next chapter deals with the question of how to treat high discretion diagnoses in a DCG model. For that analysis the DCG classification developed for the Dutch situation will be used.

6

Removing high discretion diagnoses from the Diagnostic Cost Group classification

6.1 Introduction

A DCG capitation model should not reward hospital admissions for diagnoses for which the decision to hospitalize may involve high levels of discretion. Persons with hospital admissions for high discretion diagnoses should be treated as persons without a hospitalization. In the next section high discretion is defined in the Dutch context. Section 6.3 describes the selection of high discretion diagnoses and is followed by a section that analyses the effect of removing high discretion diagnoses from the DCG classification on the predictive accuracy of the DCG model.

6.2 Definition of high discretion in the Dutch context

Ellis and Ash (1988, 1995) refined the original DCG model in a later study. They removed those diagnoses from the DCG classification for which the decision to hospitalize may involve high levels of discretion. The decision to remove a diagnosis was based on clinical judgement. Every diagnosis was judged by physicians on three dimensions of physician discretion:

- bias against non-hospital-based care: Is a hospital admission appropriate for this diagnosis?

- overstays: Is a stay at the hospital of at least three days [16] for this diagnosis necessary?
- ambiguous diagnosis: Is a diagnosis ambiguous or nonspecific and can it be used in place of more precise and accurate diagnoses?

In the capitation model DCG 0 for people who had not been hospitalized was expanded with the diagnoses that had the lowest future expenses and with so-called high discretion diagnoses. The reduction in R^2 due to the classification of high discretion diagnoses in the lowest group was about 1.4 percentage points. Van Vliet and Van de Ven (1993) applied the original DCGs developed by Ash et al. (1986, 1989) and the discretion DCGs (Ellis and Ash, 1988, 1995) to a data set with Dutch private insureds of all age-groups. For the original DCGs their results were similar to those of Ash et al. (1986, 1989). For those in DCG 0 of the discretion DCGs who are hospitalized, the average future costs were substantially above (almost five times) that of the non-hospitalized population in DCG 0. Since, moreover, the average costs for DCGs 2 to 6 were roughly the same, Van Vliet and Van de Ven concluded that the discretion DCGs in this form are not really applicable in the Dutch situation.

6.2.1 *Bias against non-hospital-based care*

For the Dutch situation bias against non-hospital-based care is the main dimension in judging whether or not the decision to hospitalize involves a high level of discretion. High discretion diagnoses are those diagnoses for which day case treatment, a form of outpatient care, may be an acceptable alternative for hospital admission. Day case treatment has become very popular during the last decade. Because of the high costs of inpatient treatment, substitution of day case treatment for hospital admissions was encouraged as a means to cost-containment. This substitution of day case treatment, also called day care or short stay, for inpatient care is important in the context of risk-adjusted capitation payments. Risk-adjusted capitation should induce sickness funds to concentrate on cost-containment and

[16] Only the diagnoses from hospital admissions of at least three days were used in the US studies on the DCG model.

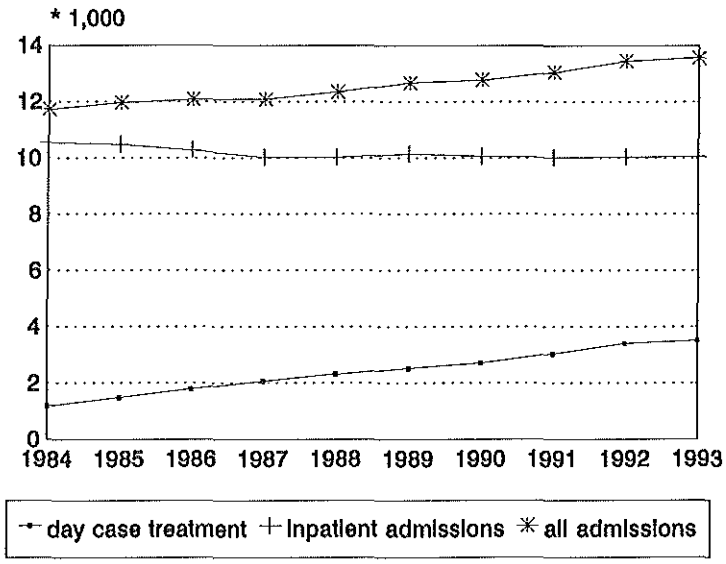
efficiency. The incentives for efficiency and cost-containment are reduced when a hospital admission of a person who could be treated in an ambulatory setting e.g. day case treatment, leads to a higher capitation payment next year for this person. The capitation should not reward hospital admissions for diagnosis appropriate for day case treatment. The next section describes day case treatment in the Netherlands in the period 1984-1993.

6.2.2 *Day case treatment in the Netherlands*

During the last decade day case treatment has become popular in the Netherlands. In 1984 only 10% of all admissions were for day case treatment, while in 1993 26% of the total number of admissions were for day case treatment (SIG, 1994). In absolute figures as well as per 100,000 inhabitants the number of day case treatments increased in the period 1984-1993. Figure 6.1 shows the total number of admissions (including day case treatments), the number of inpatient admissions and the number of day case treatments per 100,000 inhabitants over this period. The total number of admissions increased from 1984 to 1993. The average yearly growth over the last five years of this period was 1.9%. The number of inpatient admissions decreased. This reduction of admissions for inpatient care occurred mainly in the years from 1984 to 1987. Since then the number of inpatient admissions per 100,000 inhabitants remained almost the same. Over the period 1984 to 1993 the number of day case treatment increased enormously. However, the yearly growth in this period decreased. In 1985 the yearly growth of the number of day case treatments was 25%; in 1993 the yearly growth was reduced to 3%.

The proportion of day case treatments in the total number of admissions varied by age group. Children, aged 1 to 14, had most admissions for day case treatment per 100,000 inhabitants (SIG, 1988, 1990, 1991a, 1991b, 1993, 1994). Besides variation with age there is also geographical variation in the proportion day case treatment (NZI, 1990, 1994).

Figure 6.1 Number of admissions per 100,000 inhabitants



Source: NZI (National Hospital Institute)

The increase of the number of day case treatments was not accompanied by a reduction of the number of inpatient admissions. This suggests that the increase of day case treatment is not the result of substitution of day case treatment for hospital admissions. The increase in the number of day case treatments resulted in an increase of the total number of admissions. In the context of risk-adjusted capitation payments it is still important not to reward inpatient admissions for diagnoses appropriate for day case treatment. In order to achieve this, one has to know for which diagnoses day case treatment is the appropriate type of care.

In the Netherlands there are no generally accepted lists of diagnoses or procedures for which day case treatment is appropriate for medical reasons such as 'the guidelines for day case surgery', composed by the Royal Collage of Surgeons of England (1992). The National Association of

Specialists (LSV) stated that day case treatment is a valuable alternative for both inpatient and outpatient treatment of patients. However, the LSV judged composing a list with indications for which day case treatment is theoretically appropriate as less meaningful (LSV, 1984). Nevertheless, in 1985 the Sickness Fund Council (ZFR) composed a list of 51 procedures appropriate for day case treatment (ZFR, 1985). In 1993 the coordinating trade association of private health insurers (KLOZ) composed a list of 100 procedures, which are on medical grounds appropriate for day case treatment (KLOZ, 1993). These two lists partly overlap. From the 51 procedures of the ZFR only 32 occurred on the KLOZ-list. In 1993 the National Council of Public Health (NRV) in its advice on the development of day case treatment in hospitals recommended that health insurance organizations and the professional associations of medical specialist together compose lists of indications appropriate for day case treatment (NRV, 1993). Because of the lack of such lists, the selection of high discretion diagnoses in the context of risk-adjusted capitation payments e.g. diagnoses appropriate for day case treatment, is based on an empirical analysis.

6.3 The selection of high discretion diagnoses

Several lists of diagnoses appropriate for day case treatment were composed based on an empirical analysis of all hospital admissions in the Netherlands in 1993. For this analysis a database was used containing information on all approximately two million hospital admissions in 1993 in the Netherlands for both day case treatment and inpatient treatment [17]. For each admission the following variables were known: the length of stay, the age of the patient, the medical procedure performed and the diagnosis.

[17] Data source: National Medical Registration (LMR) by SIG (Information Centre for Health Care).

For the selection of diagnoses appropriate for day case treatment two methods were used. One method took the diagnosis as the starting point. The other method used the medical procedure performed as a starting point.

The method using the *diagnosis* as a starting point comprised the following steps:

- All admissions for diagnoses in the form of a three-digit ICD code that occurred at least 200 times in the data set were selected.
- For each ICD code the 'percentage day case treatment' was calculated by dividing the number of admissions for day case treatment by the total number of admissions with this ICD code.
- Each ICD code with a percentage of 50% or more day case treatment was selected for the list of diagnoses appropriate for day case treatment.

This method resulted in a list of 60 ICD codes.

The method using the *medical procedure* as a starting point comprised the following steps:

- All admissions for medical procedures (in the form of a so-called CVV-codes [18]) that occurred at least 200 times in the data set were selected.
- For each medical procedure the 'percentage day case treatment' was calculated by dividing the number of admissions for day case treatment by the total number of admissions with the same procedure code.
- Admissions with medical procedures for which the 'percentage day case treatment' was 50% or more were selected resulting in a list of 91 medical procedures.
- In the panel data set only ICD code are available and not the codes of the medical procedures performed during the admission. Therefore the ICD codes occurring most frequently together with the selected medical procedures were considered. For all selected

[18] CVV: Classificatie Voor Verrichtingen (classification for medical procedures) by SIG (Information Centre for Health Care).

procedures the percentage of the most frequently occurring ICD codes varied from 21% to 99%. Even the second and third most frequently occurring ICD codes were taken into account as long as they occurred in at least 20% of the admissions together with a medical procedure. This resulted in the selection of a list of 67 ICD codes. 37 ICD codes on this list also occurred on the list selected by the method using the ICD code as a starting point. The remaining 30 (=67-37) ICD codes formed, together with the 60 ICD codes that were already selected by the method using diagnoses as a starting point, the first list of diagnoses appropriate for day case treatment.

For 90 three-digit ICD codes day case treatment occurred in more than 50% of the admissions. The diagnoses with these ICD codes were considered high discretion diagnoses. When a discretion DCG model is used, persons with these diagnoses are placed in DCG 0 together with those without a hospital admission.

The choice to define diagnoses, for which in more than 50% of day case treatments occurred as high discretion diagnoses, is arbitrary. Therefore, for 'percentages day case treatment' of 35%, 25% and 15%, lists were also composed of high discretion diagnoses. For each of the 'percentage day case treatment' the steps described above were followed. The only difference being that the selection of diagnoses was based on a decreasing 'percentage day case treatment' resulting in an increasing number of diagnoses to be placed in DCG 0 [19]. Table 6.1 shows which percentages of the inpatient admissions in the panel data set will be placed in DCG 0 for the various lists of high discretion diagnoses. When a person has several admissions in a year, it might happen that this person is not placed into DCG 0 but into a lower DCG based on a diagnosis belonging to another admission.

When diagnoses for which the 'percentage day case treatment' is higher than 50% were placed in DCG 0 this concerns about a quarter of the admissions in the panel data set. The smaller the 'percentage day case

[19] Lamers et al. (1995) give an overview of these diagnoses.

treatment' is, the larger the number of admissions placed in DCG 0. More than half of the admissions is placed into DCG 0 when diagnoses for which the 'percentage day case treatment' is 15% or more are considered as high discretion diagnoses.

Table 6.1 Percentages of admissions appropriate for day case treatment

	% admissions [a] to DCG 0	% admissions [a] to lower DCG
diagnoses placed in DCG 0:		
diagnoses > 50% day case treatment	26 %	1 %
diagnoses > 35% day case treatment	36 %	2 %
diagnoses > 25% day case treatment	45 %	2 %
diagnoses > 15% day case treatment	55 %	2 %

[a] The presented percentages are the averages over the years 1988, 1989, 1990 and 1991.

Table 6.2 shows from which DCGs the admissions placed into DCG 0 are coming for various percentages day case treatment.

Table 6.2 Initial DCGs of admissions appropriate for day case treatment and placed in DCG 0

	Placing in DCG 0 of persons with admissions for diagnoses with:			
	> 50% day case treatment	> 35% day case treatment	> 25% day case treatment	> 15% day case treatment
Initial DCG:				
DCG 1	49.6 % [a]	48.7 % [a]	49.3 % [a]	51.3 % [a]
DCG 2	45.8 %	41.5 %	40.4 %	37.4 %
DCG 3	2.5 %	8.1 %	8.9 %	10.0 %
DCG 4	2.1 %	1.7 %	1.4 %	1.3 %
DCG 5	---	---	---	0.07 %

[a] The presented percentages are the averages over the years 1988, 1989, 1990 and 1991.

For all percentages day case treatment of about half of the persons with an admission for a high discretion diagnoses were assigned to DCG 1 before they were placed into DCG 0. When the 'percentage day case treatment' decreases, relatively fewer persons who were initially assigned to DCG 2 and relatively more persons initially assigned to DCG 3 are placed into DCG 0. When diagnoses for which the 'percentage day case treatment' is 15% or more are considered as high discretion diagnoses, even persons initially assigned to DCG 5 are placed into DCG 0 together with the persons without a hospital admission.

The mean costs in 1992, expressed as a ratio with regard to the overall mean costs, of the persons placed into DCG 0 for the various versions of high discretion diagnoses are presented in table 6.3. The mean costs in 1992 for persons placed into DCG 0 increase when the 'percentage day case treatment' to define high discretion diagnoses decreases. For the persons with a hospital admission in 1991 placed into DCG 0 the mean costs in 1992 are higher than for persons with a comparable admission in former years.

Table 6.3 Mean costs in 1992 of persons with admissions appropriate for day case treatment and placed in DCG 0, expressed as a ratio with regard to the overall mean costs

	Ratio = actual costs / mean costs in 1992			
	Placing in DCG 0 of persons with admissions for diagnoses with:			
	> 50% day case treatment	> 35% day case treatment	> 25% day case treatment	> 15% day case treatment
in 1988	1.80	2.08	2.17	2.17
in 1989	2.02	2.22	2.36	2.50
in 1990	2.07	2.16	2.58	2.69
in 1991	2.69	3.03	3.30	3.72

6.4 The predictive accuracy of the DCG model without high discretion diagnoses

This section describes the effect on the predictive accuracy of the DCG model of placing persons with a admission to the hospital for diagnoses appropriate for day case treatment into DCG 0. Table 6.4 shows the R^2-values for estimation and prediction for the one-year DCG model and various forms of the one-year discretion DCG model using the DCG classification developed for the Dutch situation (chapter 5). The one-year DCG with only persons without an admission to the hospital placed into DCG 0 serves as a reference model for the one-year discretion DCG models.

Table 6.4 R^2-values x 100 for estimation and prediction of costs in 1992

	estimation	prediction
One-year DCG		
diagnoses placed in DCG 0:		
none	8.38	6.60
diagnoses > 50% day case treatment	8.04	6.36
diagnoses > 35% day case treatment	8.00	6.55
diagnoses > 25% day case treatment	7.79	6.15
diagnoses > 15% day case treatment	6.76	5.33

Table 6.4 shows a small reduction in R^2-values for both estimation and prediction when diagnoses appropriate for day case treatment are placed into DCG 0. The R^2-values for the discretion DCG model which placed diagnoses with a 'percentage day case treatment' of 35% or more into DCG 0 hardly differ from the R^2-values of the model which placed diagnoses with a 'percentage day case treatment' of 50% or more into DCG 0. The R^2-values of these models are still almost twice the R^2-value of the demographic model (see table 5.6). The R^2-values decrease for the discretion DCG models that placed diagnoses with 'percentages day case treatment' of 25% and 15% or more into DCG 0.

Table 6.5 shows the coefficients of DCG models with DCG classifications with diagnoses appropriate for day case treatment placed in DCG 0. When admissions for more diagnoses are placed into DCG 0, i.e. a decreasing 'percentage day case treatment', the coefficient for DCG 1 decreases. In the case of placing diagnoses into DCG 0 with a 'percentage day case treatment' of at least 15%, the remaining admissions in DCG 1 do not contribute significantly in estimating costs in 1992. When a coefficient of a DCG increases when admissions for diagnoses with a decreasing 'percentage day case treatment' are placed into DCG 0, this means that from the DCG concerned admissions for diagnoses with relative lower follow-up costs are removed.

Table 6.5 Coefficients of DCGs from estimating costs in 1992 by one-year DCG models with DCG classifications with diagnoses appropriate for day case treatment placed in DCG 0

	Placing in DCG 0 of persons with admissions for diagnoses with:			
	> 50% day case treatment	> 35% day case treatment	> 25% day case treatment	> 15% day case treatment
1991 DCGs				
DCG 1	875	822	744	538 [ns]
DCG 2	3686	4007	3732	3572
DCG 3	9115	10236	10297	8682
DCG 4 - 5	9333	9314	9292	9891
diagnosis unknown	2326	2289	2267	2228

[ns] The coefficient is not statistically significant from 0 ($p > 0.05$).

Besides R^2-values cost ratios for the subgroups defined on components of medical consumption in 1988 were also determined to evaluated the predictive accuracy of the discretion DCG models. Table 6.6 shows that the cost ratios move a little away from 1 when diagnoses appropriate for day case treatment are placed into DCG 0.

Table 6.6 Ratios of actual to predicted costs in 1992 for various subgroups formed on prior use and costs for one-year DCG models with DCG classifications with diagnoses appropriate for day case treatment placed in DCG 0

	Ratio = actual costs / predicted costs in 1992				
	Placing in DCG 0 of persons with admissions for diagnoses with:				
Risk group /interval	no diagnoses in DCG 0	> 50% day case treatment	> 35% day case treatment	> 25% day case treatment	> 15% day case treatment
DCG (Ash) 1988					
0 (no admission)	0.95[ns]	0.94[ns]	0.94[ns]	0.94[ns]	0.94[ns]
1 - 2	1.13[ns]	1.15[ns]	1.15[ns]	1.15[ns]	1.16[ns]
3 - 4	1.32	1.35	1.36	1.36	1.38
5	1.56	1.59	1.60	1.61	1.67
6 - 9	2.25	2.34	2.36	2.44	2.49
Total costs in 1988					
0 Dfl.	0.77	0.76	0.76	0.76	0.75
1 - 3200 Dfl.	1.11	1.12	1.12	1.11	1.11
> 3200 Dfl.	1.59	1.63	1.65	1.65	1.70
Outpatient costs in 1988					
0 Dfl.	0.77	0.76	0.76	0.76	0.76
1 - 1000 Dfl.	1.10	1.11	1.11	1.11	1.11
> 1000 Dfl.	1.58	1.60	1.62	1.62	1.66

[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p>0.05$).

The cost ratios of the discretion DCG model placing diagnoses with a 'percentage day case treatment' of 35% or more into DCG 0 are almost the same as the ratios of the discretion model with diagnoses in DCG 0 for which day case treatment occurs for more than half of all admissions. For the subgroup with a hospital admission in 1988 in the highest DCGs the cost ratios diverge further away from 1 when diagnoses with a 'percentage

day case treatment' of 25% or more are placed into DCG 0. For the subgroups with the highest total and the highest outpatient costs in 1988, the ratios diverge from 1 when diagnoses with a 'percentage day case treatment' of 15% or more are placed in DCG 0. The differences in cost ratios for the various one-year (discretion) DCG models are small.

6.5 Conclusion

A DCG capitation model should not reward hospital admissions for diagnoses for which the decision to hospitalize may involve high levels of discretion. For the Dutch situation high discretion diagnoses may be defined as those diagnoses for which day case treatment is a possible alternative to a hospital admission. Because of the lack of (generally accepted) lists with diagnoses or procedures appropriate for day case treatment, several of such lists were composed based on an empirical analysis of all hospital admissions in the Netherlands in 1993. A person with a hospital admission for a high discretion diagnosis is treated as a person without a hospital admission.

Placing persons with a hospital admission for high discretion diagnoses into DCG 0 together with the persons without an admission affected the predictive accuracy of the one-year DCG model. However, the differences between the models were small. The predictive accuracy in terms of R^2-values decreased when diagnoses are placed into DCG 0 for which 25% or more of the patients receive day case treatment; in terms of cost ratios when diagnoses with a 'percentage day case treatment' of 15% or more were placed into DCG 0.

The selection of high discretion diagnoses used here was based on arbitrary choices. This selection procedure might result in placing the majority of admissions within a Diagnostic SubGroup (DSG) into DCG 0, while a small fraction of the admission within a DSG e.g. less than 10%, stay in the initial DCG. In practice this might lead to shifts in coding patterns

from high discretion diagnoses to non-discretion diagnoses within the same DSG. Because the purpose of this chapter was a technical analysis of the effect on the predictive accuracy of treating persons with a hospital admission for high discretion diagnosis as persons without a hospital admission, the practical implications of using discretion DCG classifications were not addressed. The conclusion from the technical analysis was that placing hospital admissions for diagnoses with a percentage of day case treatment of 35% or more into DCG 0 seems to give the best combination of the number of high discretion diagnoses and reduction of predictive accuracy. This DCG classification will be used in the next chapter as the discretion DCG classification. However, before a discretion DCG classification can be used in practice the implications should be considered such as potential shifts in coding patterns. Clinical judgement may be useful in developing such a discretion DCG classification.

7

The association between Diagnostic Cost Groups and health indicators: An indication of construct validity

7.1 Introduction

In section 2.2 requirements for ideal risk-adjusters were described. One of these requirements is validity: a risk-adjuster should measure the need for health services utilization and define a system of adjustment in which the cells are relatively homogeneous with regard to this need for health care (Van de Ven and Van Vliet, 1992; Gerritse and Poelert, 1991). A risk-adjuster should adequately adjust capitation levels for systematic differences in health status (Ash et al., 1989). Risk-adjusted capitation payments should account for predictable variations in annual per-person health care expenditures, as far as these are related to health status. According to Hornbrook and Goodman (1991) a good risk-assessment model should include measures of the beneficiary's health status. All potential risk-adjusters should relate to permanent or semi-permanent dimensions of health.

So, there are two conditions DCGs should meet to be a valid risk-adjuster. First, DCGs should have the ability to predict future health care expenditures. This refers to predictive validity or criterion validity. A measure or an instrument has predictive validity when it is able to predict a criterion (Nunnally, 1978). For DCGs future health care expenditures are the criterion. The previous chapters showed that the DCG model did a better job in predicting next year's health care expenditures than models containing demographic variables only, indicating that DCGs have predictive validity. The second condition to be met is that DCGs measure (semi-)permanent health. By (semi-)permanent health, a person's health is meant as it persists over a longer period of time e.g. for several years. This

second condition refers to construct validity. Construct validity is an evaluation of the extent to which the measurement corresponds to theoretical concepts (constructs) concerning the phenomenon under study (Last, 1988). In this chapter the construct validity of DCGs is examined. To see whether DCGs capture (semi-)permanent health, the correlations between DCGs and the health indicators from the health survey are estimated. The next section describes the method used, followed by three sections analyzing the association of DCGs with perceived health, the number of chronic diseases and other health indicators respectively.

7.2 Method

This chapter examines the association between DCGs and the health indicators from the health survey (see table 3.3). Because the object of this analysis is to find out whether DCGs measure (semi-) permanent health, the focus will be on the more general health indicators: perceived health status (the single item) and the number of chronic conditions.

Persons with one or more hospital admissions in 1992 are assigned to DCGs according to the diagnoses from these admissions. For the assignment to DCGs the DCG classification as developed in chapter 5 and the discretion DCG classification from chapter 6 are used. Chronically ill persons do not have a hospital admission every year. The probability that these persons will be assigned to a DCG increases when the base-line period to gather information on inpatient diagnostic information becomes longer. Therefore, persons are also assigned to DCGs based on their hospital admissions during the three-year period from 1990 to 1992. For people with more than one hospital admission during these three years the diagnosis classified in the highest DCG is used. The 'three-year DCGs' are expected to measure semi-permanent health better than the one-year DCGs. Table 7.1 shows the distribution of respondents to the health survey over the one-year and three-year DCGs for the two DCG classifications. Persons with a hospital admission for which the diagnosis is

missing are classified into a separate DCG: 'DCG diagnosis unknown'. In the ranking of DCGs this group is placed between DCG 1 and DCG 2 based on its coefficients from the regression analyses used in the preceding chapters relating the risk-adjusters to health care expenditures (see tables 5.10 and 6.5).

Table 7.1 shows that 83.5% of the persons had no hospital admission in the three-year period. For the discretion DCG classification 4.1% of the persons with hospital admissions for high discretion diagnoses is placed in DCG 0 together with the persons without a hospital admission.

Table 7.1 Distribution of one-year and three-year DCGs

	DCG classification		discretion DCG classification	
	one-year	three-year	one-year	three-year
DCG 0	93.4 %	83.5 %	95.3 %	87.6 %
DCG 1	2.4 %	6.5 %	1.7 %	5.1 %
Diagnosis unknown	1.4 %	3.0 %	1.4 %	3.1 %
DCG 2	1.7 %	4.4 %	0.8 %	2.2 %
DCG 3	1.0 %	2.3 %	0.7 %	1.8 %
DCG 4-5	0.1 %	0.3 %	0.1 %	0.2 %

To test whether DCG are associated with perceived health status (single item) chi-square tests are performed and adjusted standardized residuals are calculated. These adjusted standardized residuals can be used for identifying the categories responsible for a significant chi-square value. They have an approximately normal distribution with mean 0 and standard deviation 1 (Everitt, 1977). The strength of the association between DCGs and perceived health is assessed by calculating Spearman rank correlation coefficients. The Spearman correlation coefficient is used because both the DCGs and perceived health are measured on an ordinal scale.

To test the association between DCGs on the one hand and the number of chronic conditions and other health indicators on the other, analyses of

variances are performed. In these analyses the DCGs form the independent variable and the health indicators the dependent variables. For assessing the strength of the association between DCGs and the health indicators Spearman rank correlation coefficients are calculated.

7.3 The association between DCGs and health indicators

7.3.1 *Perceived health*

Table 7.2 shows per DCG the percentage of persons who perceive their health as good or very good. These percentages are presented for each of the four DCG classifications. More than 80% of the persons without a hospital admission or with an admission for a high discretion diagnosis perceive their health as (very) good. This percentage decreases with increasing DCGs. This suggests an association between DCGs and perceived health in the expected direction: the higher the DCG, the less healthy.

In table 7.2 the adjusted standardized residuals are presented between brackets. The results of the chi-square tests as well the Spearman correlation coefficients are shown in table 7.3. The association between DCGs and perceived health is statistically significant. The residuals in table 7.2 show that all DCGs contribute to this association.

The association between the three-year DCGs and perceived health is stronger than for the one-year DCGs. Placing persons with hospital admissions for high discretion diagnoses in DCG 0 reduces the strength of the association between DCGs and perceived health. The highest correlation coefficient is for the three-year DCGs, 0.22. Table 7.3 also shows the results of the chi-square tests as well as the Spearman correlation coefficients for perceived health five years ago. Perceived health five years ago and DCGs are associated. However, the relationship is weaker than for perceived health at the time of the survey.

Table 7.2 Perceived health by one-year and three-year DCGs

	One-year DCGs		Three-year DCGs	
	DCG-classification: % good (residual *)	discretion DCG-classification: % good (residual *)	DCG-classification: % good (residual *)	discretion DCG-classification: % good (residual *)
DCG 0	81.6 % (17.9)	81.1 % (15.6)	83.3 % (20.3)	82.4 % (17.8)
DCG 1	69.3 % (- 4.1)	68.8 % (- 3.6)	74.7 % (- 3.3)	73.9 % (- 3.4)
Diagnosis unknown	56.0 % (- 7.1)	56.0 % (- 7.1)	62.9 % (- 7.4)	61.9 % (- 8.0)
DCG 2	40.6 % (-12.9)	31.1 % (-11.4)	53.7 % (-14.0)	48.9 % (-11.7)
DCG 3	34.3 % (-11.3)	32.7 % (- 9.9)	41.4 % (-14.7)	40.7 % (- 13.0)
DCG 4-5	12.6 % (- 6.1)	2.1 % (-5.0)	19.4 % (- 7.9)	10.2 % (- 7.1)
Total	79.7 %	79.7 %	79.7 %	79.7 %

* The presented residuals are adjusted standardized residuals.

Health is related to age. After correcting for age the association between DCGs and perceived health is still significant. The partial correlations between DCGs and perceived health at the time of the survey as well as five years ago decrease a little with about 0.03 to 0.05.

Table 7.3 Test results for the relation between DCGs and perceived health

	Perceived health at time of participating in the survey		Perceived health five years ago	
	chi-square	rank correlation	chi-square	rank correlation
one-year DCGs:				
DCGs	413.4 *	0.18	185.4 *	0.12
discretion DCGs	324.9 *	0.15	114.8 *	0.08
three-year DCGs:				
DCGs	590.0 *	0.22	333.8 *	0.16
discretion DCGs	463.1 *	0.18	247.7 *	0.13

* All results in this table are statistically significant ($p < .01$). degrees of freedom = 5.

7.3.2 *The number of chronic diseases*

Table 7.4. presents per DCG the mean number of 'chronic diseases for which people are still under treatment' for the various DCG classifications. As expected, the number of chronic diseases still being treated increases with increasing DCG. Compared with the one-year DCGs the mean number of diseases for the three-year DCGs are smaller. Between brackets table 7.4 shows the percentage of persons per DCG without chronic diseases for which they are still under treatment. For DCG 0 more than 60% of the people are without such chronic diseases. These percentages decrease to about 4 to 6% for the highest DCGs of the one-year DCGs and to about 11 to 14% for the highest DCGs of the three-year DCGs.

Table 7.4 Mean number of treated chronic diseases by one-year and three-year DCGs

	One-year DCGs		Three-year DCGs	
	DCG-classification: mean number (% 0)	discretion DCG-classification: mean number (% 0)	DCG-classification: mean number (% 0)	discretion DCG-classification: mean number (% 0)
DCG 0	0.62 (63.0 %)	0.64 (62.4 %)	0.58 (64.8 %)	0.60 (63.8 %)
DCG 1	0.96 (45.9 %)	0.94 (49.2 %)	0.85 (53.5 %)	0.87 (54.5 %)
Diagnosis unknown	1.60 (27.9 %)	1.60 (27.9 %)	1.21 (37.3 %)	1.25 (36.4 %)
DCG 2	1.55 (29.2 %)	1.68 (26.2 %)	1.27 (38.3 %)	1.35 (35.4 %)
DCG 3	1.82 (19.3 %)	1.95 (18.2 %)	1.60 (25.7 %)	1.62 (26.9 %)
DCG 4-5	2.37 (5.7 %)	2.48 (3.9 %)	2.16 (14.0 %)	2.17 (11.4 %)
Total	.67 (61.0 %)	0.67 (61.0 %)	0.67 (61.0 %)	0.67 (61.0 %)

Table 7.5 shows the test results, F-statistics and Spearman correlation coefficients, for the association between DCGs and the number of chronic conditions. The tests are performed for the number of chronic diseases for which people are still treated and for the number of diseases regardless of treatment separately.

Table 7.5 Test results for the relation between DCGs and the number of chronic diseases

	Still treated chronic conditions		Chronic conditions regardless of treatment	
	F-statistic	rank correlation	F-statistic	rank correlation
one-year DCGs:				
DCGs	80.5 *	0.17	80.6 *	0.17
discretion DCGs	63.7 *	0.14	60.0 *	0.14
three-year DCGs:				
DCGs	106.8 *	0.20	129.3 *	0.22
discretion DCGs	82.1 *	0.17	94.8 *	0.19

All results in this table are statistically significant ($p < .01$).
* degrees of freedom = 5

Table 7.5 shows that DCGs are associated with the number of chronic diseases. The higher the DCG, the more chronic diseases. This association is stronger for the DCG classification with only persons without a hospital admission in DCG 0 than for the discretion DCGs. The association with the number of chronic diseases is stronger for the three-year DCGs than for the one-year DCGs. For the one-year DCGs it makes no difference whether the chronic diseases are still treated or not. For the three-year DCGs the association with the number of chronic diseases is stronger for the number of diseases regardless of treatment. After correcting for age the associations between DCGs and the number of chronic diseases is still statistically significant. Compared to the correlation between DCGs and the number of chronic conditions without correcting for age, the partial correlation coefficients are about 0.04 to 0.05 lower.

7.3.3 *Other health indicators*

The previous sections were concerned with the relation between DCGs and more general health indicators whereas this section deals with the association between DCGs and the more specific health indicators. These health indicators are functional disabilities (in communication and mobility), disabilities in Activities for Daily Living (ADL), psychological unwell-being, a scale measuring perceived or subjective health status and the body-mass index. Table 7.6 presents the test results, F-statistics and Spearman correlation coefficients, for the association between DCGs and these health indicators.

Table 7.6 Test results for the relation between DCGs and other health indicators

	one-year DCGs: DCG classification (discretion DCGs)		*three-year DCGs*: DCG classification (discretion DCGs)	
	F-statistic	rank correlation	F-statistic	rank correlation
Perceived health status (scale)	32.7 * (24.4 *)	0.12 (0.10)	53.1 * (36.7 *)	0.14 (0.11)
Functional disabilities	60.8 * (39.5 *)	0.14 (0.12)	93.1 * (65.4 *)	0.17 (0.14)
Disabilities in Activities for Daily Living	25.9 * (21.5 *)	0.11 (0.08)	39.2 * (34.9 *)	0.14 (0.12)
Psychological unwell-being	12.4 * (7.1 *)	0.07 (0.05)	13.2 * (6.9 *)	0.06 (0.04)
Body-Mass index	4.9 * (3.4 *)	0.05 (0.04)	11.4 * (9.4 *)	0.07 (0.06)

All results in this table are statistically significant ($p < .01$).

* degrees of freedom = 5.

The associations between DCGs and the health indicators in table 7.6 are statistically significant. DCGs show the strongest association with functional disabilities, followed by perceived health (scale) and disabilities in ADL. The DCGs have the weakest association with psychological unwell-being and the body-mass index. The associations between the discretion DCGs and the health indicators are weaker than the associations with the DCGs with only non-hospitalized people in DCG 0. Except for psychological unwell-being, the relations of the three-year DCGs are stronger than for the one-year DCGs.

After correcting for age the association between DCGs and body-mass index is no longer statistically significant. The association between DCGs and psychological unwell-being is not affected by a correction for age. Correcting for age has the strongest impact on the association between DCGs and functional disabilities. Although still statistically significant, the partial correlation is about 0.06 to 0.08 lower than the correlation uncorrected for age. For perceived health status and disabilities in ADL the reduction in correlation is about 0.03 to 0.05. The partial correlation coefficients for the association of these health indicators and DCGs are still statistically significant.

7.4 Conclusion

DCGs were statistically significant associated with the health indicators from the health survey. The relation was as expected: the higher the DCG, the less healthy. DCGs showed the strongest relation with the more general health indicators i.e. perceived health and the number of chronic diseases, followed by functional disabilities. Correction for age somewhat weakened the associations, which remained significant however. These associations between DCGs and health indicators showed that DCGs have construct validity. DCGs seem to capture (semi-)permanent dimensions of health.

The association between the health indicators and 'DCGs with only persons without a hospital admission in DCG 0' was stronger than with the discretion DCGs. The three-year DCG showed stronger associations with all the health indicators than the one-year DCGs. Although the associations with the health indicators were statistically significant, the highest correlation coefficient was only 0.22 for the three-year DCGs and perceived health and the number of chronic diseases (regardless of treatment). An explanation for this relatively low correlation might be that even with three-year DCGs the group of persons without a hospital admission (DCG 0) is large. There exist differences in health among these people. These differences are captured by the survey health indicators, but not by the DCGs.

Because the discretion DCGs showed weaker associations with the health indicators than the DCGs with DCG 0 restricted for persons without a hospital admission, and because this discretion DCG classification is the result of a technical analysis and cannot be used in practice in this form (see section 6.5), in the next chapters the discretion DCG classification will not be used. This chapter showed that three-year DCGs have a stronger association with all health indicators. The next chapter focuses on the question of whether three-year DCGs are better predictors of future health care expenditures than the one-year DCGs.

8

From a one-year DCG model to a multi-year DCG model

8.1 Introduction

Only about 6.5% of the sickness fund members has one or more hospital admissions in a year. Every year there will be persons in poor health with predictably high expenditures, who are not hospitalized and thus not classified into a DCG. In this chapter the usefulness of incorporating DCGs over *a longer base-line period* in the capitation model is examined. The one-year DCG model is extended to a two-year and a three-year DCG model. The rationale for this is twofold. First, having had a serious hospitalization in a given year might induce predictably above-average expenditures not only in the directly following year but, to a diminishing degree, also in years thereafter (without necessarily resulting in a new hospitalization). Secondly, by giving higher capitation payments for people who have been hospitalized for certain diagnoses during one of the previous years (instead of only during the last year), we increase the probability that a sickness fund will receive an appropriate capitation payment for chronically ill patients among its members. According to Gruenberg et al. (1989) a three-year period is the most appropriate to sort out the systematic and random elements of hospital use. In section 8.2 the predictive accuracy of the two-year and three-year DCG model is described.

An individual's prior year's costs is the best single predictor of his future medical expenses (see section 2.3). To see whether prior year costs are still predictive after taking into account prior inpatient morbidity, both the demographic and DCG models were expanded with the variable 'high

prior costs'. Section 8.3 describes the effect of adding this variable on the predictive accuracy of the models.

8.2 The predictive accuracy of the multi-year DCG model

The one-year DCG model is extended with dummies for the 1990 DCGs to get the two-year DCG model. The three-year DCG model comprised dummies for DCGs in 1989, 1990 and 1991 (see also table 3.4). So, the latter model captures the time that elapses between the year in which an admission takes place and the prediction period -which is 1992 in this analysis- as well as situations where people are hospitalized in more than one year of the base period. As this may provide incentives for gaming an alternative model was also investigated. In this alternative model people were classified according to their highest DCG in the three-year base period as in the 'three-year DCGs' in chapter 7. This removes the attractiveness of hospitalizing people in several years. An explorative analysis showed that the predictive performance (R^2-value) of the latter model was substantially lower than that of the three-year DCG model.

Table 8.1 presents the R^2-values for estimation and prediction for the two-year and three-year DCG model. For comparison the R^2-values of the one-year DCG model are also presented. Table 8.1. shows that each additional year of DCGs gives a small improvement compared with the one-year DCG model.

Table 8.1 R^2 x 100 for estimation and prediction of costs in 1992 for the one-year, two-year and three-year DCG model

	estimation	prediction
One-year DCG	8.38	6.60
Two-year DCG	9.04	7.17
Three-year DCG	9.40	7.83

Table 8.2 shows the estimated coefficients of DCGs by the three-year DCG model. The DCGs of all three years show the increasing trend in coefficients: the higher the DCG, the higher the coefficient. The coefficients decrease when the period between the admission and the prediction year 1992 becomes longer. This means that an admission in 1989 contributes less to estimating the costs in 1992 than an admission in the same DCG in 1991.

Table 8.2 Coefficients of DCGs from estimating costs in 1992 by three-year DCG model

	DCGs in 1991	DCGs in 1990	DCGs in 1989
DCG 1	807	774	726
DCG 2	3075	1211	1223
DCG 3	8313	2145	1787
DCG 4 - 5	8463	5766	3816
diagnosis unknown	1982	1262	--

Table 8.3 presents the cost ratios for 1992 for various subgroups defined on components of health care costs and use in 1988 (note that none of the models incorporate information on costs or use in 1988). Table 8.3 shows that, for each additional year of DCGs, the cost ratio's converge further to 1. For people with a hospital admission in 1988 for diagnoses classified in DCG 5 the actual costs in 1992 are 56% above the predicted costs of the one-year DCG model; for persons in the highest DCGs in 1988 the actual costs are 125% above the predicted costs. These differences between actual and predicted costs diminish when costs are predicted with a multi-year DCG model. The above-mentioned differences between actual and predicted costs reduce to 36% respectively 74% when costs are predicted by the three-year DCG model. For the group of persons with the highest total costs in 1988 the expected loss in 1992 reduces by 17 percentage points when costs are predicted by the three-year DCG model compared with the one-year DCG model. The difference between actual and predicted costs

for persons with the highest outpatient costs in 1988 diminishes from 58% (one-year DCG model) to 44% when costs are predicted with the three-year DCG model.

Table 8.3 Ratios of actual to predicted costs in 1992 for various subgroups formed on prior use and costs for the one-year, two-year and three-year DCG model

	Ratio = actual / predicted costs in 1992		
Risk group / interval	one-year DCG	two-year DCG	three-year DCG
DCG (Ash) in 1988			
0 (no admission)	0.95[ns]	0.95[ns]	0.96[ns]
1-2	1.13[ns]	1.11[ns]	1.09[ns]
3-4	1.32	1.26	1.19
5	1.56	1.44	1.36
6-9	2.25	1.95	1.74
Total costs in 1988			
0 Dfl.	0.77	0.78	0.80
1-3200 Dfl.	1.11	1.10	1.09
> 3200 Dfl.	1.59	1.50	1.42
Outpatient costs in 1988			
0 Dfl.	0.77	0.79	0.81
1-1000 Dfl.	1.10	1.09	1.08[ns]
> 1000 Dfl.	1.58	1.51	1.44

[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

The difference between actual and predicted costs by the demographic model (table 4.6) for persons with a hospital admission for the highest DCGs in 1988 is 209%; for persons with the highest costs in 1988 88% and for people with the highest outpatients costs in 1988 79%. Compared to these differences between actual and predicted costs the reduction of the

expected loss when costs are predicted by the three-year DCG model is even more striking.

8.3 High prior costs as an additional risk-adjuster for the DCG models

When inpatient diagnostic information is incorporated in the capitation model the question arises whether other administrative information is available that a sickness fund might use to game the capitation system. Information with predictive power for future costs that might be useful as a vehicle for cream skimming. Prior costs are good predictors of future health care expenditures. Information on prior costs is available in the administration of a sickness fund. To see to what extent prior year costs are still predictive after taking into account prior inpatient morbidity, in the form of DCGs, both the demographic and DCG capitation models were expanded with a variable 'high prior costs'. With the inclusion of prior costs in a capitation model there is always the danger of rewarding inefficient providers and encouraging more utilization than is strictly necessary (see also section 2.3). Therefore, only *high* prior costs, above a certain threshold, are used in the capitation models. Costs above the 99 percentile of the empirical distribution of costs are considered high costs. This yielded threshold values for 1989, 1990 and 1991 of 12,759, 11,488 and 17,250 guilders, respectively. The threshold values are not increasing over the years because of the incomplete data on costs in these years (see table 3.2). The costs above the thresholds are incorporated in the capitation models as continuous variables.

The 'demographic plus high prior costs' model includes the risk factors of the demographic model (see table 3.4) extended with a continuous variable for the costs above the threshold in 1991. The 'one-year DCG plus high prior costs' model is the one-year DCG model extended with the 1991 high costs variable. For the 'two-year DCG plus high prior costs' model, the two-year DCG model is extended with the high costs in 1991 and 1990, whereas the 'three-year DCG plus high prior costs' model comprised

a continuous variable for the high costs in 1989 as well. Table 8.4 shows the R^2-values for estimation and prediction of the 'plus high prior costs' variants of the demographic and the DCG models. For all models the extension with high prior costs increases the R^2-values. The R^2-values for the 'demographic plus high prior costs' model are even higher than the R^2 of the one-year DCG model (table 8.1). The highest R^2-values are for the 'three-year plus high prior costs' model.

Table 8.4 R^2 x 100 for estimation and prediction of costs in 1992 for the 'demographic plus high prior costs' and the 'one-year, two-year and three-year DCG plus high prior costs' models

	estimation	prediction
Demographic plus high prior costs	8.50	9.34
One-year DCG plus high prior costs	10.78	10.54
Two-year DCG plus high prior costs	11.65	11.33
Three-year DCG plus high prior costs	12.15	11.85

For the subgroups defined on components of health care costs and use in 1988 costs ratios for 1992 are calculated for the models with high prior costs (table 8.5). The difference in cost ratios between a particular model and the same model 'plus high prior costs' becomes clear in the cost ratios for the subgroups with the highest costs in 1988 and the groups with a hospital admission for diagnoses in DCG 5 or higher. The cost ratios for the one-year DCG model (table 8.3) are closer to 1 than the cost ratios for the 'demographic plus high prior costs' model.

Table 8.5 Ratios of actual to predicted costs in 1992 for various subgroups formed on prior use and costs for the 'demographic plus high prior costs' and 'one-year, two-year and three-year DCG plus high prior costs' models

	Ratio = actual / predicted costs in 1992			
Risk group /interval	demo-graphic + high prior costs	one-year DCG + high prior costs	two-year DCG + high prior costs	three-year DCG + high prior costs
DCG (Ash) in 1988				
0 (no admission)	0.94	0.95[ns]	0.96[ns]	0.96[ns]
1-2	1.22	1.16[ns]	1.13[ns]	1.11[ns]
3-4	1.42	1.31	1.24	1.17
5	1.72	1.52	1.41	1.31
6-9	2.56	2.16	1.84	1.63
Total costs in 1988				
0 Dfl.	0.75	0.77	0.79	0.81
1-3200 Dfl.	1.12	1.11	1.10	1.09
> 3200 Dfl.	1.73	1.56	1.46	1.36
Outpatient costs in 1988				
0 Dfl.	0.75	0.78	0.79	0.81
1-1000 Dfl.	1.11	1.10	1.09	1.08
> 1000 Dfl.	1.67	1.54	1.48	1.40

[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

8.4 Conclusion

In this chapter the performance as a risk-adjuster of DCGs defined over periods of several years was examined. The predictive accuracy, in terms of R^2-values and cost ratios, improved when DCGs over a longer period are incorporated in the model. The differences between the actual and

predicted costs in 1992 for persons with a hospital admission in the highest DCGs in 1988 diminished from 125% (one-year DCG model) to 74% when costs are predicted with the three-year DCG model. The expected loss in 1992 for people with the highest costs in 1988 decreased from 59% (one-year DCG model) to 42% (three-year DCG model). When the demographic and the DCG models were extended with information on high prior costs the predictive accuracy, especially in terms of R^2-values, further improved. The differences in cost ratios between a particular model and the same model 'plus high prior costs' was noticed in the cost ratios for the subgroups with the highest costs and the highest DCGs in 1988.

To evaluate the predictive accuracy of the models we compared their R^2-values. We also looked at the predictive accuracy more directly, namely by assessing how well they predict costs for various relevant subgroups in the data set. Although the difference in R^2-value between a one-year and three-year DCG model was small, the reduction in expected loss in 1992 was substantial: 29% for the persons with the highest costs in 1988 and 41% for those with a hospital admission in the highest DCGs in 1988. The R^2-value of the 'demographic plus high prior costs' model was higher than the R^2-value of the one-year DCG model. However, the expected losses in 1992 for the latter model were smaller when we looked at the cost ratios. For the evaluation of the predictive accuracy of various models it is useful to take both R^2-values and cost ratios into account.

With the inclusion of prior costs in a capitation model there is always the danger of rewarding inefficient sickness funds and encouraging more utilization than is strictly necessary. Therefore, we used only *high* prior costs, above a certain threshold. Only one percent of the sickness fund members had high costs, exceeding the threshold, in one year. The estimated models indicated that for every guilder spent above the threshold a sickness fund would receive only 31 cents via next year's capitation payments; for every guilder spent in 1990 15 cents and for every guilder spent in 1989 8 cents via the 1992 capitation payments. Therefore, it is not likely that high prior costs as a risk-adjuster will provide incentives for inefficiency.

9

An evaluation of the predictive accuracy of DCG models using health survey data

9.1 Introduction

In one year about 6.5% of the sickness fund members of all ages are hospitalized; in a three-year period about 16%. Chapter 8 showed that the three-year DCG model explained about 9% of the differences between individuals in health care expenditures; the 'three-year plus high prior costs' model about 12%. The maximum predictable portion of medical expenditure variation is estimated at about 20% (Newhouse et al., 1989; Van Vliet, 1992). The fact that the three-year DCG model captures only one-half of the predictable variance implies either that there are groups of people with predictable high expenditures who have had no hospital admissions in the three-year period or that the DCG classification needs further refinements.

Besides (diagnostic information from) prior utilization, self-reported health measures are predictors of future health care expenditures (section 2.5). Health status indicators like perceived health, functional health status and chronic conditions improved the predictive accuracy of demographic capitation models (Thomas and Lichtenstein, 1986; Van Vliet and Van de Ven, 1992; Hornbrook and Goodman, 1995, 1996). Perceived health status and functional health status measures and self-reported chronic conditions are mostly obtained by surveys. In general, health survey information is not routinely collected by sickness funds and is not available in the sickness fund administrative data. Thus, risk-adjusters based on survey information are at present inappropriate in the Dutch context. However, health survey information can be used to evaluate the predictive accuracy of capitation models. In this chapter the predictive accuracy of DCG

models is evaluated using health survey data. The next section describes the specific methods used for this evaluation. In section 9.3 subgroups based on health survey data are identified for which the 'three-year DCG plus high prior costs' model is unable to predict costs accurately. In section 9.4 the effect on the predictive accuracy is studied of extending various capitation models from the previous chapters with the relevant survey information.

9.2 Method

In this chapter the demographic model and DCG (plus high prior costs) models are estimated by means of ordinary least squares, with an individual's annual health care expenditures in 1994 as dependent variable and the various sets of risk factors as independent variables (N=52,674). The estimated coefficients are used to predict costs for those who completed the health survey. On the basis of the survey answers various subgroups are formed. For these subgroups predicted costs are compared with actual costs in 1994 by means of cost ratios.

After identifying those subgroups for which the actual costs in 1994 are significantly higher than the costs predicted by the 'three-year DCG plus high prior costs' model, we studied which subset of relevant survey variables can improve the predictive accuracy of this DCG model. For the selection of survey variables stepwise regression procedures were used. Each capitation model is extended with dummy's for the selected survey variables.

For the assessment of the predictive accuracy of the demographic, the one-year and three-year DCG model and the variants of these models extended with survey information the data set is restricted to respondents to the health survey only (N=10,570). Because of the relatively small size of this data set the split-sample method is repeated 30 times. The ability of alternative capitation models to predict future costs is evaluated in the

prediction data sets by means of R^2-values and cost ratios for the subgroups based on medical consumption in 1988.

Furthermore, the predicted costs of various models are compared. For each comparison predicted costs of a 'capitation' model are compared with the costs predicted by a more comprehensive 'selection' model. A 'good risk' is defined as somebody whose capitation payment i.e. the costs predicted by the capitation model, is higher than the predicted costs of the selection model. For a 'bad risk' the opposite holds. A positive difference between capitation payment minus predicted costs of the selection model concerned implies a predictable profit; a negative difference a predictable loss. The predictable profits and losses based on diagnostic information, high prior costs and survey information are assessed.

9.3 A comparison of actual and predicted costs for subgroups formed with health survey data

The models are estimated on the total data set. Not all persons in the data set were eligible for the health survey. First, there is a group of persons older than 90 years of age or receiving institutional care who were excluded from the health survey. A second group is formed by persons whose administrative data were added to the data set in the second stage, when the health survey was already conducted. There is also a group of persons who where eligible for the health survey, but did not receive a questionnaire. Finally, persons who received a questionnaire can be divided in respondents and non-respondents. For all these groups the actual costs in 1994 were, statistically, not significantly different from the predicted costs by the various capitation models. This indicates that the results are not affected by the study design used for the health survey.

For the group of respondents the actual and predicted costs in 1994 did not differ. However, subgroups within the group of respondents can be identified for which actual costs differ significantly from the predictions of

the capitation models. In this study the primary interest is in those subgroups whose actual costs are higher than the costs predicted by the models. The results for these subgroups will be presented in this section.

For subgroups based on the health survey, cost ratios are calculated. For some of the subgroups based on health survey information the actual costs did not differ significantly from the costs predicted by the demographic model ($p > .05$). Table 9.1 shows the cost ratios for the demographic model for these subgroups. For one subgroup in table 9.1, contrary to the expectations, actual costs in 1994 are significantly lower than the predicted costs of the demographic model. This concerns 'respondents and/or his parents born outside the Netherlands'. The demographic model appears to predict accurately the costs of persons in subgroups based on demographic variables like education and marital status.

Table 9.2 shows the cost ratios for subgroups for which the actual costs are higher than the predictions of the demographic model but not significantly different from the predicted costs by the 'three-year DCG plus high prior costs' model. Apparently health care expenditures in the subgroups of table 9.2 can be predicted by the DCG models.
The differences between actual and predicted costs diminish when costs are predicted by DCG models compared to the demographic models. The more sophisticated the model, the smaller the differences between actual and predicted costs. In table 9.2 the cost ratios are shown for the least sophisticated DCG model for which, in actual costs, there is no longer any statistically significant difference from the predicted costs of the DCG model concerned. For example, the costs of persons still treated for asthma, chronic bronchitis and COPD can be predicted by the one-year DCG model. The actual costs of this subgroup are not statistically significantly different from the predicted costs of the two-year and three-year DCG model and the variants of the DCG models extended with high prior costs as well. On the other hand, the 'three-year DCG plus high prior costs' model is the only model, which can predict accurately the costs of persons still treated for a serious disease of the kidney and for persons using home help during the last year.

Table 9.1 Ratios of actual to predicted costs in 1994 for various subgroups based on the health survey for which there is no statistically significant difference between the actual costs and predictions of the demographic model

Subgroup [a]	%	Ratio = actual costs in 1994/ predicted costs in 1994	
		no risk-adjustment[b]	demographic
Use of drugs without prescription during last 2 weeks	21.3 %	0.90	0.93[ns]
Use of speech therapy during last year	1.6 %	0.77	1.21[ns]
Use of social work during last year	2.7 %	1.60	1.44[ns]
Use of ambulatory mental care during last 5 years	4.3 %	1.05	1.01[ns]
Use of dentures	22.6 %	1.91	1.03[ns]
Visit a dentist last year	4.7 %	1.19	1.01[ns]
High propensity for health care use	7.6 %	1.53	1.17[ns]
Body-mass index ≥ 30 (Obesitas)	7.4 %	1.15	0.94[ns]
Low education	26.1 %	1.49	1.00[ns]
Being widowed	7.0 %	2.46	1.07[ns]
Respondent and/or his parents born outside the Netherlands	7.6 %	0.67	0.69
Living alone	13.3 %	1.63	1.05[ns]

[a] For subgroups of persons with various chronic conditions (still under treatment) and persons using various prescribed drugs the difference between actual costs and costs predicted by the demographic model were not statistically significant. This concerned the following groups: persons using the following prescribed drugs: aspirins, medicine for common cold, vitamins, skin agents, medicines for allergy and hormones and persons still treated for perinasal, frontal or maxillary sinusitis, disease of the liver and liver cirrhosis, chronic cystitis, chronic spinal disorder and slipped disc, arthritis of hands or feet, other rheumatoid arthritis (for longer than 3 months), epilepsy, migraine, serious skin disease, depression and other nervous conditions, stones in the kidney, hypertension, stroke and effect of stroke, stomach and duodenal ulcer, gall-stones and inflammation of the gall-bladder and prolapse.

[b] Mean costs in 1994 (1430 guilders) for respondents to the health survey are used as predicted costs.

Table 9.2 Ratios of actual to predicted costs in 1994 of the model for which actual and predicted costs are no longer significantly different for various subgroups based on the health survey

Subgroup	%	Ratio = actual costs in 1994 / predicted costs in 1994			
		no risk-adjustment[a]	demographic	DCG model (actual = predicted)	
≥ 8 days in bed due to illness during the last 6 months	6.6 %	1.62	1.52	1.21[ns]	one-year
≥ 5 health problems (perceived health questionnaire) [b]	37.7 %	1.51	1.18	1.10[ns]	two-year
Psychological unwell-being [b]	9.1 %	1.58	1.26	1.18[ns]	one-year
Still treated for:					
asthma chronic bronchitis COPD	5.0 %	2.04	1.57	1.36[ns]	one-year
serious heart disease/heart attack	1.8 %	2.91	1.43	1.18[ns]	one-year
disorder of large or small bowel	1.8 %	2.33	1.56	1.31[ns]	one-year
serious disease of the kidney	0.2 %	13.60	9.78	1.77[ns]	3-y+hpc[c]
thyroid disorder or goitre [b]	1.2 %	2.97	1.95	1.64[ns]	one-year
arthrosis of knees, hips hands [b]	7.1 %	2.25	1.28	1.28[ns]	one-year
Contact with specialist last year	37.0 %	1.56	1.28	1.11[ns]	two-year
Expect contact with specialist next year	29.3 %	1.76	1.36	1.11[ns]	3-year
Hospital admission last year	9.6 %	2.43	1.84	1.00[ns]	two-year
Expect hospital admission next year	4.5 %	2.47	1.99	1.25[ns]	one-year
Use home help during last year[b]	3.0 %	3.86	1.61	1.24[ns]	3-y+hpc[c]
Use of prescribed drugs:					
drugs heart diseases, hypertension	10.6 %	2.47	1.23	1.17[ns]	one-year
diuretics	5.1 %	2.99	1.39	1.17[ns]	two-year
laxatives	2.2 %	3.20	1.56	1.35[ns]	one-year
digestants	4.5 %	2.23	1.40	1.20[ns]	one-year
antibiotics	4.0 %	2.01	1.68	1.40[ns]	one-year
medicine for Asthma	2.7 %	2.42	1.86	1.55[ns]	two-year
medicine for the eyes	3.3 %	2.62	1.44	1.30[ns]	3-year

[a] Mean costs 1994 (1430 guilders) of respondents to the survey are used as predicted costs.
[b] Adults only.
[c] 3-y+hpc = 'three-year DCG plus high prior costs' model.
[ns] No statistically significant difference between actual and predicted costs (t-test, $p > 0.05$).

For some subgroups the actual costs are significantly higher than the predictions of the 'three-year DCG plus high prior costs' model. The cost ratios for these groups are presented in table 9.3. Table 9.3 show that the differences between actual and predicted costs diminish when costs are predicted by DCG models compared to the demographic models. For example, for those with poor perceived health the difference between actual and predicted costs of 41% (demographic model) reduces to about 30% and 20% when costs are predicted by the one-year and three-year DCG models respectively. For some subgroups in table 9.3, like persons still treated for cancer and persons using an alternative practitioner, the extension of the DCG models with high prior costs does not lead to a further reduction of the difference between actual and predicted costs. All groups in table 9.3 have actual costs that cannot be predicted accurately by any of the capitation models. Even the 'three-year DCG plus high prior costs' model gives predicted costs that are significantly lower than the actual costs.

9.4 The predictive accuracy of DCG models extended with survey information

9.4.1 *The selection of survey information improving the predictive accuracy of the DCG models*

It is likely that the subgroups in table 9.3 overlap to a certain extent. Therefore, it was studied which subset of survey variables in table 9.3 could improve the predictive accuracy of the 'three-year DCG plus high prior costs' model. The survey variable 'using medicine for diabetes mellitus' was excluded from this analysis, because of the overlap with the variable 'still under treatment for diabetes mellitus'. A subset of eight variables was selected on the basis of stepwise regression procedures [20].

[20] Three methods for model selection were used: maximum R^2, maximum R^2-improvement and Mallows CP selection method. All methods resulted in the same set of eight survey variables.

Table 9.3 Ratios of actual to predicted costs in 1994 for various subgroups based on the health survey for which actual costs are statistically significantly higher than the predictions of the 'three-year DCG plus high prior costs' model

		Ratio = actual costs in 1994 / predicted costs in 1994						
Subgroup	%	no risk-adjustment [a]	demographic	demographic + HPC	one-year DCG	one-year DCG + HPC	three-year DCG	3-year DCG + HPC
Poor perceived health (single item)	20.1 %	2.16	1.41	1.35	1.30	1.29	1.20	1.20
Poor perceived health 5 years ago	17.9 %	1.90	1.36	1.29	1.26	1.24	1.17	1.16
≥ 3 chronic conditions	17.7 %	2.20	1.54	1.44	1.39	1.37	1.26	1.25
Diabetes mellitus still treated	1.7 %	3.71	1.90	1.77	1.70	1.67	1.51	1.52
Cancer still under treatment	1.2 %	3.55	2.18	2.11	1.91	1.98	1.43	1.52
Functional disabilities (mobility and communication) [b]	9.6 %	3.01	1.49	1.41	1.38	1.35	1.30	1.27
Problems with activities daily living[b]	3.5 %	4.33	1.93	1.75	1.65	1.62	1.51	1.48
≥4 GP visits last 2 months	4.1 %	2.64	1.77	1.63	1.50	1.49	1.33	1.34
Use of physiotherapy last year	17.5 %	1.63	1.37	1.31	1.25	1.24	1.17	1.16
Use of home nursing last year	2.3 %	3.39	2.06	1.70	1.85	1.65	1.55	1.40
Use of alternative practitioner last year	7.6 %	1.45	1.42	1.45	1.39	1.42	1.38	1.43
Use of ≥ 5 prescribed drugs during the last 2 weeks	3.3 %	3.74	1.95	1.63	1.68	1.53	1.48	1.39
Use of sedatives / tranquillizers	7.4 %	2.36	1.33	1.27	1.26	1.23	1.22	1.19
Use medicine rheumatoid arthritis	4.6 %	2.71	1.58	1.53	1.41	1.44	1.37	1.43
Use medicine for diabetes mellitus	1.8 %	3.80	1.84	1.73	1.58	1.59	1.41	1.45

[a] Mean costs in 1994 (1430 guilders) for respondents to the health survey are used as predicted costs.

[b] Adults only.

The selected variables were: perceived health (2 categories), having functional disabilities (2 categories), consulting the GP (3 categories), use of home nursing (2 categories), number of prescribed drugs used (5 categories), having cancer in combination with yes/no under treatment (3 categories), having diabetes in combination with yes/no under treatment (3 categories) and the use of medicine for rheumatoid arthritis (2 categories).

9.4.2 *R^2-values and cost ratios*

The demographic, the one-year DCG and the three-year DCG model as well as variants of these models 'plus high prior costs' were extended with 14 dummy's for the eight selected survey variables. Table 9.4 shows the R^2-values for estimation and prediction of costs in 1994. The R^2-values for estimation are higher than for prediction. The ranking of the models according to increasing R^2-values is, for estimation and prediction, almost the same. The R^2-values for estimation and prediction of costs in 1994 are a little higher than the R^2-values for estimation and prediction of costs in 1992 as presented in the previous chapters. The largest differences occur for the models including 'high prior costs'.

With the demographic model almost 4% of the variance in costs in 1994 among individuals can be predicted. The predictive accuracy of the demographic model, in terms of R^2-values, improves when the survey variables are included in the model. The one-year and three-year DCG models give comparable results. The models extended with survey variables yield higher R^2-values. The R^2-value for the 'demographic plus survey' model is about the same as for the one-year DCG model, a demographic model extended with last year's DCGs. The models extended with high prior costs yield the highest R^2-values. Extending a model with both survey information and high prior costs increases the R^2-values compared to the models extended with high prior costs only.

Table 9.4 R^2-values * 100 for estimation and prediction for various capitation models

	R^2 * 100 [a]	
	estimation	prediction
Model:		
Demographic model	5.90	3.78
+ high prior costs	12.04	11.22
+ survey	9.31	6.00
+ survey + high prior costs	14.73	12.55
One-year DCG model	8.79	6.48
+ high prior costs	13.28	12.01
+ survey	11.46	7.89
+ high prior costs + survey	15.60	12.99
Three-year DCG model	11.07	8.00
+ high prior costs	15.75	12.70
+ survey	13.02	8.64
+ high prior costs + survey	17.48	13.11

[a] The R^2-values are the means over 30 estimations of the models.

Cost ratios for the year 1994 are calculated for various subgroups defined on components of health care costs and utilization in 1988. Table 9.5 shows the ratios for the demographic capitation models; table 9.6 for the one-year DCG models and table 9.7 for the three-year DCG models. To define subgroups based on DCGs in 1988 the DCG classification developed for the Dutch situation (see chapter 5) is used. In the former chapters these groups were defined using the original DCG classification of Ash et al. (1986, 1989). In this chapter costs in 1994 are predicted. In 1992 and 1993 people left the sickness fund. Because of this loss the subgroup of persons with a hospital admission in DCG 6-9 (according to the DCG classification of Ash et al.) became very small. Therefore, the DCG classification developed for the Dutch situation is used.

Table 9.5 Ratios of actual to predicted costs in 1994 for various subgroups formed on prior use and costs for demographic models

	Ratio = actual costs in 1994 / predicted costs in 1994				
Risk factor / interval	no risk-adjustment [a]	demographic	+ high prior costs	+ survey	+ HPC [b] + survey
DCG [c] in 1988					
0 (no admission)	0.89	0.90	0.92	0.93	0.94
1	1.44	1.43	1.28	1.23	1.14
2	2.91	1.88	1.75	1.54	1.49
3 - 5	5.30	3.15	2.09	2.18	1.69
Total costs in 1988					
0 Dfl.	0.68	0.79	0.82	0.91	0.92
1 - 3200 Dfl.	1.16	1.04	1.05	0.97	0.98[ns]
> 3200 Dfl.	2.65	1.93	1.61	1.54	1.37
Outpatient costs 1988					
0 Dfl.	0.69	0.80	0.83	0.91	0.93
1 - 1000 Dfl.	1.17	1.06	1.05	0.99[ns]	0.98[ns]
> 1000 Dfl.	2.31	1.64	1.52	1.34	1.28

[a] In this case predicted costs is the average in 1994 of 30 estimation data sets of respondents to the survey: 1381 Dutch guilders.
[b] HPC = high prior costs.
[c] The DCG classification developed for the Dutch situation (chapter 5) is used.
[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

Table 9.5 shows that for people in DCG 3-5 (in 1988) the average actual costs in 1994 are 215% above the costs predicted by the demographic model. For people in DCG 0 (no hospital admission) the actual costs are 10% below the costs predicted by the demographic model. The difference between actual and predicted costs diminishes for all subgroups when costs are predicted by a demographic model that is extended with high prior costs or survey variables. The cost ratios for the 'demographic plus survey' model are closer to one than those for the 'demographic plus high prior costs' model. Especially for the groups with the highest outpatient costs in

1988 the differences between actual and predicted costs reduces from 64% when costs are predicted by the demographic model to 34% when costs are predicted by this 'demographic plus survey' model. Extending the demographic model with both high prior costs and survey information leads to predictions closest to the actual costs.

Table 9.6 Ratios of actual to predicted costs in 1994 for various subgroups formed on prior use and costs for one-year DCG models

	Ratio = actual in 1994 / predicted costs in 1994			
Risk group /interval	one-year DCG	+ high prior costs	+ survey	+ high prior costs + survey
DCG [a] in 1988				
0 (no admission)	0.92	0.93	0.94	0.95
1	1.32	1.24	1.18	1.12
2	1.67	1.64	1.44	1.44
3 - 5	2.40	1.90	1.90	1.61
Total costs in 1988				
0 Dfl.	0.83	0.84	0.93	0.94
1-3200 Dfl.	1.03[ns]	1.03	0.97	0.97[ns]
> 3200 Dfl.	1.68	1.51	1.43	1.32
Outpatient costs in 1988				
0 Dfl.	0.83	0.85	0.93	0.94
1-1000 Dfl.	1.04	1.04	0.98[ns]	0.98[ns]
> 1000 Dfl.	1.48	1.43	1.27	1.25

[a] The DCG classification developed for the Dutch situation (chapter 5) is used.
[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

Table 9.6 and 9.7 show comparable results for the one-year DCG models and the three-year DCG models respectively. The predictions of the DCG models extended with high prior costs deviate less from actual costs than the predictions of the models without high prior costs. Extending the

DCG models with survey information reduces the difference between actual and predicted costs even more. The 'DCG plus high prior costs plus survey' models gives ratios closest to 1.

Table 9.7 Ratios of actual to predicted costs in 1994 for various subgroups formed on prior use and costs for three-year DCG models

	Ratio = actual in 1994 / predicted costs in 1994			
Risk group /interval	three-year DCG	+ high prior costs	+ survey	+ high prior costs + survey
DCG [a] in 1988				
0 (no admission)	0.93	0.94	0.94	0.95
1	1.24	1.17	1.15	1.09[ns]
2	1.50	1.47	1.37	1.36
3 - 5	2.00	1.64	1.73	1.48
Total costs in 1988				
0 Dfl.	0.86	0.87	0.94	0.95
1-3200 Dfl.	1.02[ns]	1.03[ns]	0.97	0.98[ns]
> 3200 Dfl.	1.50	1.36	1.35	1.25
Outpatient costs in 1988				
0 Dfl.	0.86	0.88	0.94	0.95
1-1000 Dfl.	1.03[ns]	0.98[ns]	0.98[ns]	0.97[ns]
> 1000 Dfl.	1.39	1.37	1.25	1.23

[a] The DCG classification developed for the Dutch situation (chapter 5) is used.
[ns] No statistically significant difference between actual and predicted mean costs (two-sided t-test, $p > 0.05$).

The cost ratios for the 'demographic plus survey' model are closer to 1 than the cost ratios of the one-year DCG model. A demographic model extended with both survey variables and one-year DCGs further diminishes the differences between actual and predicted costs. The cost ratios of this 'one-year DCG plus survey' model are closer to 1 than the ratios of the three-year DCG model.

The most comprehensive model, the 'three-year DCG plus high prior costs plus survey' model, gives the best predictions. For persons with a hospital admission in the highest DCGs in 1988 the actual costs are 215% above the predictions of the demographic model and only 48% more than the predictions of this most comprehensive model; for the groups with the highest costs in 1988 the difference between actual and predicted costs drops from 93% (demographic model) to 25% when costs are predicted with the most comprehensive model. For the 7% of members with the highest outpatient costs in 1988 the predictions of the 'three-year DCG plus survey' model and the 'three-year DCG plus high prior costs plus survey' model are almost the same. The actual costs are about 25% above these predictions.

9.4.3 *Predictable profits and losses*

The predicted costs of various models are compared. For each comparison predicted costs of a 'capitation' model are compared with the costs predicted by a more comprehensive 'selection' model. A 'good risk' is defined as somebody whose capitation payment i.e. the costs predicted by the capitation model, is higher than the predicted costs of the selection model. For a 'bad risk' the opposite holds. Currently the capitation payments in the Netherlands are based on the demographic model. Table 9.8 shows the percentages of good and bad risks using the demographic model as the capitation model and the DCG models and the models including high prior costs as selection models. This table also includes the differences between capitation payment minus predicted costs of the selection model concerned. A positive difference implies a predictable profit; a negative difference a predictable loss. Within the groups of good and bad risks a distinction between persons with and without a hospital admission in the preceding year (1993) is made.

Table 9.8 shows that using the demographic model as the capitation model about one third of the members is a bad risk based on high costs in the preceding year with a mean predictable loss of 265 guilders. The mean predictable profit for the good risks is 129 guilders. Based on diagnostic information from hospital admissions in the preceding year 10% of the

members is identified as a bad risk with a mean predictable loss of 1,700 Dutch guilders. When diagnostic information from hospital admissions in the three preceding years is available 15% of the members form the group of bad risks with a mean predictable loss of almost 1,900 guilders. Within the groups of bad risks the mean predictable loss for persons with a hospital admission in the preceding year is much higher than for persons without a hospital admission. Based on the 'demographic plus high prior costs' model and both one-year DCG models the mean predictable loss for bad risks without a hospital admission in the preceding year is 64 guilders or less. Even the 95th percentile of the predictable losses for this group is less than 100 guilders. The mean predictable loss for bad risks with an admission to the hospital in the preceding year varies from 2800 to about 3400 guilders. Based on diagnostic information from the three preceding years, whether or not in combination with information about high prior costs, the mean predictable loss of bad risks without a hospital admission in the preceding year is about one third of the mean predictable loss of bad risks with a hospital admission. The percentiles presented in table 9.8 show that the distributions of predictable losses are very skewed.

The predicted costs of the demographic and the DCG (plus high prior costs) models as capitation models are compared with the variants of these models with survey information as selection models. The predictable profits and losses based on survey information are presented in table 9.9. All models identify almost the same percentage of the persons as bad risks. The percentage of bad risks varies from 28% to 32%. Table 9.9 shows that the more comprehensive the model, the lower the predictable profits and losses based on survey information. The predictable profits reduced from 434 guilders per member (demographic model) to 322 guilders ('three-year DCG plus high prior costs' model); the predictable losses from 1082 to 713 guilders per member. A distinction is made within the groups of good and bad risks between members with and without a hospital admission in the preceding year. Among the good risks relatively many persons with a hospital admission occur. Their predictable profits are higher than for those without an admission, varying from 538 guilders ('demographic plus high prior costs' model as capitation model) to 750 guilders (one-year DCG model as capitation model). The group of bad risks contains many persons

Table 9.8 Capitation payment (demographic model) in 1994 minus predicted costs by selection models for good and bad risks by hospital admission in 1993 [a]

Capitation model	Selection model		Capitation payment minus predicted costs: %	Capitation payment minus predicted costs: costs	Hospital admission in 1993	Capitation payment minus predicted costs by selection model: %	mean	75th percentile	95th percentile
Demographic	Demographic	good risks	67.3 %	129	no	93.9 %	126	166	454
					yes	6.1 %	187	291	557
	+ high prior costs	bad risks	32.7 %	- 265	no	93.4 %	- 42	- 41	- 80
					yes	6.6 %	- 3399	- 3389	-18264
Demographic	One-year DCG	good risks	90.0 %	192	no	99.8 %	192	237	581
					yes	0.2 %	185	263	499
		bad risks	10.0 %	-1717	no	38.9 %	- 19	- 25	- 52
					yes	61.1 %	- 2797	- 3631	- 7636
Demographic	One-year DCG	good risks	86.0 %	208	no	99.1 %	207	248	728
					yes	0.9 %	347	512	916
	+ high prior costs	bad risks	14.0 %	-1271	no	60.6 %	- 64	- 41	- 82
					yes	39.4 %	- 3128	- 3350	- 10933

Table 9.8 Continued

Demo-graphic	Three-year DCG	good risks	84.8 %	338	no	99.2 %	337	407	982
					yes	0.8 %	464	566	1120
		bad risks	15.2 %	- 1872	no	63.3 %	- 1178	- 1640	- 3968
					yes	36.7 %	- 3067	- 3935	- 8718
Demo-graphic	Three-year DCG + high prior costs	good risks	84.8 %	330	no	98.6 %	326	380	1064
					yes	1.4 %	570	770	1468
		bad risks	15.2 %	-1854	no	66.3 %	- 1062	- 1328	- 3682
					yes	33.7 %	- 3413	- 3600	- 11918

[a] Good risks: costs predicted by capitation model (demographic model) > costs predicted by selection model.
Bad risks: costs predicted by capitation model (demographic model) ≤ costs predicted by selection model.

Table 9.9 Predictable profits and losses in 1994 based on health survey information as a supplement to various capitation models for good and bad risks by hospital admission in 1993 [a]

Capitation model	Selection model		Capitationpay-ment minus predicted costs		Hospital admission in 1993	Capitation payment minus predicted costs by selection model			
			%			%	mean	75th percentile	95th percentile
Demo-graphic	Demo-graphic	good risks	71.0 %	434	no	95.3 %	428	532	1232
					yes	4.7 %	559	747	1545
	+ survey	bad risks	29.0 %	- 1082	no	89.9 %	- 1023	- 1471	- 3442
					yes	10.1 %	- 1604	-2300	- 4988
Demo-graphic	Demo-graphic	good risks	69.1 %	400	no	95.1 %	393	493	1092
					yes	4.9 %	538	737	1520
+ high	+ HPC [b]	bad risks	30.9 %	- 909	no	90.6 %	- 862	-1198	- 3010
prior costs	+ survey				yes	9.4 %	- 1357	-1931	- 4248
One-year DCG	One-year DCG	good risks	71.8 %	373	no	94.3 %	350	437	1017
					yes	5.7 %	750	1000	1792
	+ survey	bad risks	28.2 %	- 964	no	92.4 %	- 923	- 1326	- 3141
					yes	7.6 %	- 1464	- 2148	- 4381

Table 9.9 Continued

One-year DCG + high prior costs	One-year DCG + HPC [b] + survey	good risks	69.7 %	364	no	94.2 %	343	430	958
					yes	5.8 %	699	938	1678
		bad risks	30.3 %	- 853	no	97.7 %	- 818	- 1137	- 2890
					yes	7.3 %	- 1302	- 1885	- 3942
Three-year DCG	Three-year DCG + survey	good risks	70.6 %	325	no	94.3 %	306	378	930
					yes	5.7 %	645	876	1648
		bad risks	29.4 %	- 796	no	92.4 %	- 760	- 1084	- 2668
					yes	7.6 %	- 1226	- 1773	- 3714
Three-year DCG + high prior costs	Three-year DCG + HPC [b] + survey	good risks	68.5 %	322	no	94.2 %	303	379	890
					yes	5.8 %	614	830	1576
		bad risks	31.5 %	- 713	no	92.7 %	- 681	- 936	- 2455
					yes	7.3 %	- 1117	- 1603	- 3385

[a] Good risks: costs predicted by capitation model (model without survey information) > costs predicted by selection model (same model extended with survey information).
Bad risks: costs predicted by capitation model (model without survey information) ≤ costs predicted by selection model (same model extended with survey information).

[b] HPC = high prior costs.

without a hospital admission in the preceding year. Their predictable losses vary from 1023 guilders (demographic model as capitation model) to 681 guilders ('three-year DCG plus high prior costs' model as capitation model). The percentiles of the predictable losses based on survey information show that the distributions of the predictable losses are skewed.

9.5 Conclusion

The currently used demographic capitation model has poor predictive accuracy. The previous chapters showed that the predictive accuracy improved when the demographic model is extended with diagnostic information from prior hospitalizations. When the demographic and the DCG models were extended with information on high prior costs the predictive accuracy further improved. These conclusions were supported again in this chapter by an analysis of predictable profits and losses under a demographic capitation model. Based on high costs in the preceding year about one third of the members were a bad risk with a mean predictable loss of 265 guilders; based on diagnostic information from hospital admissions in the preceding year 10% of the members formed the group of bad risks with a mean predictable loss of 1,700 guilders; and when diagnostic information from hospital admissions in the three preceding years is available 15% of the members were a bad risk with a mean predictable loss of almost 1,900 guilders. Within the groups of bad risks the mean predictable loss for persons with a hospital admission in the preceding year was much higher than for persons without a hospital admission. The mean predictable loss for this group varied from 2800 to about 3400 guilders.

Information about health and medical consumption from a health survey was used to form subgroups. For most of these subgroups actual costs in 1994 no longer differed from the predicted costs when costs were predicted by a DCG model. However, for some groups the difference was still

statistically significant even when the costs were predicted with the 'three-year DCG plus high prior costs' model. A subset of eight survey variables improved the predictive accuracy of this model.

The predictive accuracy of the demographic model improved when the model is extended with the selected survey variables. This is consistent with other research (Van Vliet and Van de Ven, 1992; Hornbrook and Goodman, 1995, 1996; Gruenberg et al., 1996). A study by the Physician Payment Review Commission (1994b) showed that the predictive accuracy of a model containing demographic variables and Ambulatory Diagnostic Groups could be improved by including survey measures of functional health status and chronic conditions. This study showed comparable results for the DCG model. In terms of R^2-values, the predictive accuracy of the 'demographic plus survey' model and the one-year DCG model were comparable. Extending the one-year DCG with survey information improved the predictive accuracy, both in terms of R^2-values and cost ratios.

The more extensive the model, the lower the predictable profits and losses based on survey information. However, the mean predictable losses were still substantial for the three-year DCG model and 'three-year DCG plus high prior costs' model. This suggest that the DCGs and the survey information are, to a certain extent, complementary in their ability to predict future health care expenditures. Most persons who could be considered a bad risk based on health survey information were not hospitalized in the preceding year. Even when costs were predicted with the 'three-year DCG plus high prior costs' model the mean predictable losses for this group were almost 700 guilders. To improve the capitation system the set of risk-adjusters should be extended with a measure which captures the predictable high costs of this group.

10

Discussion and conclusion

Adequate risk-adjustment is critical to the success of market-oriented health care reforms. Risk-adjusted capitation payments are intended to provide the competing sickness funds with an incentive for efficiency instead of indulging in risk selection. The capitation payments should be adjusted for the need of health care of the members. Risk-adjusted capitation payments should account for predictable variations in annual per-person health care expenditures, as far as these are related to health status. Demographic factors, the most commonly used risk-adjusters, are relatively poor predictors of future health care utilization of individuals. Risk-adjustment based only on demographic factors is inadequate because it does not adjust sufficiently for health status. The capitation system can be improved by extending the set of risk-adjusters with factors that are more directly related to health. Proposed health status risk-adjusters for refining the capitation formula are measures of prior costs and prior use (whether or not in combination with diagnostic information), perceived health status and functional health status (Lubitz, 1987; Epstein and Cumella, 1988). The present study focused on diagnostic information, because this potential risk-adjuster 'seems to strike the best balance of practicality, accuracy and appropriate incentives' (Kronick et al., 1996). At the start of this study most sickness funds in the Netherlands recorded diagnostic information from hospitalizations on a regular basis. Therefore, inpatient morbidity models were examined. The DCG model developed by Ash et al. (1989) was used as a starting point for the empirical analyses. The present study showed that this DCG model, developed with US data of people of 65 years and over, could be successfully applied to Dutch sickness fund data of persons of all ages.

In this study a DCG classification for the Dutch situation has been developed with Dutch sickness fund data of both aged and non-aged

persons. The resulting clustering of ICD codes into five DCGs was based on follow-up costs in the second and third year after hospital admission. From a theoretical point of view the clustering based on follow-up costs in the second and third year after admission was preferred. By not using the costs in the first year after hospital admission, follow-up costs which are related to self-limiting acute conditions were excluded. The resulting DCG classification is based on health problems that are likely to be chronic.

The predictive accuracy of the demographic model improved when it was extended with diagnostic information from prior hospitalizations. The R^2-value of the one-year DCG model with the DCG classification developed for the Dutch situation was twice that of the demographic model. The increase in predictive accuracy was also found when actual costs and costs predicted by various capitation models were compared for subgroups formed on medical consumption and health care costs in the past. For sickness fund members with a hospital admission in the highest DCGs in 1988 the average costs in 1992 were 209% above the costs predicted by the demographic model. This difference between actual and predicted costs reduced to 125% when costs were predicted by the one-year DCG model. For the 6% group of persons with the highest costs in 1988 the expected loss in 1992 reduced by 30% when predicted costs were based on the one-year DCG model compared with the demographic model. For the 6% group of persons with the highest outpatient costs in 1988 the reduction in expected loss was 24% when predicted costs were based on the DCG model compared with the demographic model. This study also showed that besides predictive *validity,* DCGs have construct validity. DCGs seem to capture (semi-)permanent dimensions of health.

Validity is one of the desirable properties of risk-adjusters. Other requirements a risk-adjuster ideally should meet are: absence of perverse incentives, invulnerability to manipulation and obtainability or feasibility. Models incorporating diagnostic information from prior hospitalizations may create *perverse incentives*, i.e. inappropriately hospitalizing people in order to obtain a higher capitation payment for them in the future. Assuming that no individuals would be hospitalized without any medical problem, the extent of this problem depends on a sickness fund's present

marginal costs and its future marginal benefits of treating individuals on either an inpatient or outpatient basis. This problem could be reduced by a good monitoring system and by not rewarding admissions for high discretion diagnoses. For the Dutch situation high discretion diagnoses may be defined as those diagnoses for which day case treatment is a possible alternative for a hospital admission. Based on an empirical analysis of all hospital admissions in the Netherlands in 1993 several lists of high discretion diagnoses were composed. A diagnosis occurred on such a lists when more than a threshold percentage of all patients received day case treatment for the diagnosis concerned. Weiner et al. (1996) defined their 'Hospital Dominant' marker in a similar way (see section 2.4.2). Placing persons with a hospital admission for high discretion diagnoses into DCG 0 together with the persons without an admission resulted in a slight reduction of the predictive accuracy of the one-year DCG model.

In principle there may be a possibility of *manipulation*, namely by 'inflating' diagnoses to move hospitalizations to better paid categories. 'DCG creep' might occur when prospective capitation payments are based on DCGs. This may be prevented to a large extent by employing fewer groups, by monitoring and by putting alike diagnoses in the same diagnostic group.

In a situation where diagnostic information is recorded on a regular basis, a capitation system (partly) based on these diagnoses seems administrative *feasible*. In 1992, at the start of this study, diagnostic information from hospitalizations of their members was in principle available to sickness funds at an individual level. In 1997 this situation has been changed. However, once the weights for DCGs are determined, sickness funds can provide the Central Fund with aggregated information about the number of their members per DCG. A demonstration project in the United States showed that the DCG payment system is operationally feasible (Nelson and Arnold, 1990). In 1993 California embraced a broad range of underwriting reforms in the small-group insurance market. In 1996 twenty four health plans participated in the health insurance plan of California (HIPC). The legislation allowed, but did not mandate, use of a prospective risk-adjustment process to minimize differences in the actuarial risk of the

health plans participating in the cooperative. The HIPC adopted a prospective method to measure and adjust for the risk-based differences in health plan enrolment. The case mix is based on demographic factors and diagnostic information (Shewry et al., 1996).

A capitation system using the DCG model is operationally feasible, but more complex than a system using a demographic model. Implementing a DCG capitation model will improve the predictive accuracy, but it will also result in more intensive monitoring activities to prevent diagnostic up-coding and inappropriate hospitalizations. Risk-adjustment of capitation payments is a dynamic process of continuously improving and updating the system.

The predictive accuracy of a three-year DCG model, a demographic model extended with DCGs for the three preceding years, is higher than that of a one-year DCG. However, with regard to feasibility the three-year DCG model will be more complex. A three-year DCG model requires that sickness funds provide the Central Fund with diagnostic information about three years. Sickness funds have an annual open enrolment period. For persons changing sickness funds the diagnostic information from hospitalizations for the preceding three years is not available. Therefore, it is recommended to start with implementing the one-year DCG model. As long as generally accepted lists of high discretion diagnoses, i.e. appropriate for day case treatment, are lacking one can consider to reward only admissions for diagnoses in the highest DCGs (DCG 3 to DCG 5) with a higher capitation payment.

The one-year DCG model explained 8% of the variations in health care expenditures, where the maximum predictable portion of medical expenditure variation is estimated at about 20% (Newhouse et al., 1989; Van Vliet, 1992). This suggests that the one-year DCG model is unable to remove fully the potential for risk selection. The following solutions to this problem were examined: a longer base-line period for gathering information on inpatient morbidity and a model based on diagnostic information as well as prior costs. When DCGs over a longer period were incorporated in the model the predictive accuracy improved, in terms of

both R^2-values and cost ratios, i.e. actual divided by predicted costs for various subgroups of high and low risk individuals. For 10% of the sickness fund members the costs predicted by the one-year DCG model were higher than the predictions of the demographic model. So in the case of demographic risk-adjusters these 10% would be predictably 'bad risks' based on diagnostic information from hospital admissions in the preceding year. The mean predictable loss for these bad risks was 1,700 guilders. For 15% of the persons the costs predicted by a three-year DCG model were higher than the costs predicted by a model with only demographic risk-adjusters. Based on diagnostic information from hospital admissions in the three preceding years, the mean predictable loss of these 15% was almost 1,900 guilders. Within the groups of bad risks the mean predictable loss for persons with a hospital admission in the preceding year was much higher than for persons without a hospital admission. The mean predictable loss for the group with an admission to the hospital varied between 2800 and 3100 guilders. These results show that with their administrative information on diagnoses sickness funds can easily identify groups of members with predictable losses in case of demographic capitation payments.

Prior costs is a good predictor of future health care expenditures. Information on prior costs is available in the administration of a sickness fund. To see to what extent prior year costs are still predictive after taking into account prior inpatient morbidity, DCG capitation models were expanded with 'high prior costs'. With the inclusion of prior costs in a capitation model there is always the danger of rewarding inefficient sickness funds and encouraging more utilization than is strictly necessary. Therefore, only *high* prior costs, above a certain threshold, were used in the capitation models. When the demographic and the DCG models were extended with information on high prior costs the predictive accuracy, especially in terms of R^2-values, further improved. For the subgroups with the highest costs and the highest DCGs in 1988 the differences between actual and predicted costs reduced when a capitation model was extended with high prior costs in comparison to the same model without this extension. As the elasticity of high prior costs in the 'one-year DCG plus

high prior cost' model is around 0.3, it is not likely that high prior costs as a risk-adjuster will provide incentives for inefficiency.

Self-reported health measures mostly obtained with surveys are also predictors of future health care expenditures. In general, health survey information is not routinely collected by sickness funds and is not available in the sickness fund administrative data. Thus, risk-adjusters based on survey information are at present inappropriate in the Dutch context. In this study health survey information was used to evaluate the predictive accuracy of DCG capitation models. A subset of eight survey variables improved the predictive accuracy of the most comprehensive model, the 'three-year DCG plus high prior costs' model.
For the demographic model, the one-year and three-year DCG models and the variants of these models extended with high prior costs, the predicted costs were compared to the predicted costs of the same model extended with the relevant survey information. For about 30% of the persons their predicted costs based on a model extended with survey information were higher than the predicted costs of the same model without survey information. The average predictable losses of these bad risks varied from 1082 guilders per person (demographic model) to 713 guilders ('three-year DCG plus high prior costs' model). The more extensive the model, the lower the predictable profits and losses based on survey information. However, even for the three-year DCG models the mean predictable losses were still substantial. This suggests that the DCGs and the survey information are to a certain extent complementary in their ability to predict future health care expenditures. The group of bad risks based on survey information contained many persons without a hospital admission in the preceding year. These persons were assigned to DCG 0 (no admission), so their medical expenditures were predicted with only demographic variables. These persons have poor perceived health, are more likely to have certain functional disabilities, frequently visit their GP, use relatively many prescribed drugs and receive relatively more often home nursing. Their health care expenditures are higher than for healthy persons with the same demographic characteristics. A risk-adjuster based on outpatient morbidity or prescribed drugs is expected to improve the capitation systems for these persons. Since ambulatory diagnoses are not routinely recorded by most

Dutch sickness funds and information on prescribed drugs becomes more and more available on the individual level, a risk-adjuster based on prescribed drugs like the revised CDS developed by Clark and coworkers (1995) may be a promising option for further improvement of the risk-adjusted capitation system for Dutch sickness funds.

The health care expenditures used in this study included the costs of all hospital services, both inpatient and outpatient specialist care, dental care, obstetrics and maternity care, paramedical services and sick-transport. Since 1996, the basic benefit package of the sickness fund insurance has been extended with outpatient prescription drugs and technical aids. In the same year the hospital costs has been split into production-dependent and production-independent costs. The sickness funds are financially responsible for the production-dependent hospital costs only, because they can influence these costs by negotiating with hospitals about the hospital production for their members. Lamers (1996) compared the predictive accuracy of the demographic and DCG capitation models for expenditures including the production-independent hospital costs and excluding the costs for prescribed drugs and technical aids (as used in the empirical analyses in this study) with the predictive accuracy for expenditures excluding the production-independent hospital costs and including the costs for prescribed drugs and technical aids (the situation since 1996). The predictability of the latter type of expenditures is higher than that of the former. This holds for all the capitation models. Even when sickness funds have no financial responsibility for the production-independent hospital costs, by using information on prior utilization they can identify groups of members with predictable losses. These losses reduce when the currently used demographic model is extended with DCGs. So, the results of the present study are still relevant after the extension of the basic benefit package of the sickness fund insurance with outpatient prescription drugs and technical aids and the introduction of retrospective reimbursement for the production-independent hospital costs.

The predictive accuracy of the best risk-adjustment models that have been developed in various empirical studies is not perfect. Therefore, some have argued that partial capitation, also called mixed or blended payments,

should be preferred to fully prospective payments. Such mixed payment systems balance incentives for efficiency in the delivery of health care services and cost containment on the one hand and incentives for risk selection on the other hand (Newhouse, 1996). As long as sickness funds can identify high and low costs members within a risk cell and they are able to discriminate effectively between these groups resulting in substantial profits, sickness funds have incentives for cream skimming. Although even the best risk-adjustment models that have been developed leave much of the variation in expenditures unexplained, this is only a problem if sickness funds are able to both acquire more detailed information on a patient's health status than is included in the risk models, and act selectively on this information (Luft, 1995). This is easier to accomplish in the case of a demographic capitation model than with a DCG model. The incentives for cream skimming will be reduced when the payments to the sickness funds are only partly based on a prospective risk-adjustment model. In the Netherlands a partial capitation system and a system of excess of loss is used for this reason. However, retrospective reimbursement [21] of a part of the costs incurred by their members not only reduces the incentives for cream skimming, it also reduces the incentives for efficiency and cost containment for sickness funds and may be unfair to efficient sickness funds. Extending the set of demographic risk-adjusters with DCGs reduces the incentives for cream skimming while the incentives for efficiency and cost containment remain.

With the administrative information on prior utilization sickness funds can easily identify groups of members whose expenditures are expected to be far above their demographic capitation payments. The predictable losses for these high cost members reduce when the capitation payments are based on DCGs as well. Though not 'perfect', DCG models can predict a substantial amount of the predictable variance. There is no need to wait

[21] Several forms of retrospective payments can be used in addition to prospective capitation payments, for example, establishing an outlier pool to allow health plans to recover part or all of their costs for very high cost cases (Beebe, 1992). Van Barneveld et al. (1996) studied three main variants of mandatory pooling i.e. high risk pooling, proportional pooling and excess of loss pooling, which is similar to an outlier pool. Combinations of various forms of retrospective payments are also possible.

until the perfect model is available to refine the capitation payment formula. The use of diagnostic information from prior hospitalizations seems a promising option to start with improving the formula.

References

Anderson, G.P., E.P. Steinberg, N.R. Powe, et al., 1990, Setting payment rates for capitated systems: A comparison of various alternatives. Inquiry, 27, 225-233.

Anderson, G., E.P. Steinberg, J. Whittle et al., 1990, Development of clinical and economic prognoses from Medicare claims data. JAMA, 263 (7), 967-972.

Ash, A., F. Porell, L. Gruenberg, et al., 1986, An analysis of alternative AAPCC models using data from the continuous Medicare History file. Brandeis/Boston University, Boston.

Ash, A., F. Porell, L. Gruenberg, et al., 1989, Adjusting Medicare capitation payments using prior hospitalization data. Health Care Financing Review, 10 (4), 17-29.

Beck, K., P. Zweifel, 1996, Improving on risk adjustment in Swiss health insurance. Paper presented at the International Health Economics Association Inauguration Conference, Vancouver, Canada, May 19-23, 1996.

Beebe, J.C., 1988, Medicare reimbursement and regression to the mean. Health Care Financing Review, 9 (3), 9-22.

Beebe, J.C., 1992, An outlier pool for Medicare HMO payments. Health Care Financing Review, 14, 59-63.

Berk, M.L., A.C. Monheit, 1992, The concentration of health expenditures: an update. Health Affairs, 11, 145-149.

Bowen, B., 1995, The practice of risk adjustment. Inquiry, 32, 33-40

Brown, R.S, D.G. Clement, J.W. Hill, et al., 1993, Do Health Maintenance Organizations work for Medicare? Health Care Financing Review, 15 (1), 7-23.

CBS (Netherlands Central Bureau of Statistics), 1993, De onverzekerden tegen ziektekosten 1985-1992; een actualisering (The uninsured for health care costs 1985-1992: an updating). Maandbericht Gezondheidsstatistiek, 12 (10), 9-10.

References

Chinitz D.P., 1994, Reforming the Israeli health care market. Social Science & Medicine, 39 (10), 1447-1457.

Clark D.O., M. von Korff, K. Saunders, et al., 1995, A chronic disease score with empirically derived weights. Medical Care, 33 (8), 783-795.

Culler, S.D., CM Callahan, F.D. Wolinsky, 1995, Predicting hospital costs among older decedents over time. Medical Care, 33 (11), 1089-1105.

Den Dulk, C.J., H. van der Stadt, J.M. Vliegen, 1992, Een nieuwe maatstaf voor stedelijkheid: de omgevingsadressendichtheid (A new measure for degree of urbanization: the address density of the surrounding area). Maandbericht Bevolking (CBS), 7, 14-27.

Dillman, D.W., 1978, Mail and telephone surveys: the total design method. Wiley, New York.

Dowd, B., R. Feldman, I. Moscovice, et al., 1996, An analysis of selectivity bias in the Medicare AAPCC. Health Care Financing Review, 17 (3), 35-57.

Duan, N., W.G. Manning, C.N. Morris, J.P. Newhouse, 1983, A comparison of alternative models for the demand for medical care. Journal of Business and Economic Statistics, 1 (2), 115-126.

Ellis, R.P., A. Ash, 1988, Refining the diagnostic cost group model: a proposed modification of the AAPCC for HMO reimbursement. Boston University, Boston.

Ellis, R.P., A. Ash, 1995, Refinements to the diagnostic cost group (DCG) model. Inquiry, 32, 418-429.

Ellis, R.P., G.C. Pope, L.I. Iezonni, et al., 1996, Diagnosis-based risk adjustment for Medicare capitation payments. Health Care Financing Review, 17 (3), 101-128.

Enthoven, A.C., 1978, Consumer-choice of health plan: a national-health-insurance proposal based on regulated competition in the private sector. New England Journal of Medicine, 298, 709-720.

Epstein, A.M., E.J. Cumella, 1988, Capitation payment: using predictors of medical utilization to adjust rates. Health Care Financing Review, 10 (1), 51-69.

Everitt, B.S., 1977, The analysis of contingency tables. Chapman and Hall, London.

Everitt, B.S., 1993, Cluster Analysis (third edition). Edward Arnold, London.

Farley, D.O, G.M. Carter, J.D. Kallich, et al., 1996, Modified capitation and treatment incentives for end stage renal disease. Health Care Financing Review, 17 (3), 129-142.

Files, A., M. Murray, 1995, German risk structure compensation: enhancing equity and effectivenes. Inquiry, 32 (3), 300-309.

GAO (US General Accounting Office), 1994a, Health care reform: Considerations for risk adjustment under community rating. GAO/HEHS-94-173, Washington D.C.

GAO (US General Accounting Office), 1994b, Medicare: Changes to HMO rate setting method are needed to reduce program costs. GAO/HEHS-94-119, Washington D.C.

Gauthier, A.K., J.A. Lamphere, N.L. Barrand, 1995, Risk selection in the heatlh care market: a workshop overview. Inquiry, 32 (1), 14-22.

Gerritse, R., J.D. Poelert, 1991, Middelen op maat, naar een verdeelstelsel voor uitkeringen aan zorgverzekeraars (Towards a capitation payment system for health insurers). Instituut voor Onderzoek van Overheidsuitgaven, Den Haag.

Glennerster, H., M. Matsaganis, 1993, The UK health reforms: The fundholding experiment. Health Policy, 23, 179-191

Graf von der Schulenburg, J.M., 1994, Forming and reforming the market for third-party purchasing of health care: a German perspective. Social Science & Medicine, 39 (10), 1473-1481.

Gruenberg, L., E. Kaganova, M.C. Hornbrook, 1996, Improving the AAPCC with health-status measures from the MCBS. Health Care Financing Review, 17 (3), 59-75.

Gruenberg, L., C. Tompkins, F. Porell, 1989, The health status and utilization patterns of the elderly: implications for setting Medicare payments to HMOs. In: Scheffler, M.C. and L.F. Rossiter (eds), Advances in health economics and health services research, 10, JAI Press inc., London, 41-73.

Hellinger, F.J., 1995, Selection bias in HMOs and PPOs: a review of the evidence. Inquiry, 32 (2), 135-142.

Hornbrook, M.C., M.J. Goodman, 1991, Health plan case mix: definition, measurement, and use. In: Hornbrook, M.C. (ed), Advances in health economics and health services research, 12, JAI Press inc., London, 111-148.

Hornbrook, M.C., M.J. Goodman, 1995, Assessing relative health plan risk with the RAND-36 health survey. Inquiry, 32 (1), 56-74.

Hornbrook, M.C., M.J. Goodman, 1996, Chronic disease, functional health status, and demographics: a multi-dimensional approach to risk adjustment. Health Services Research, 31 (3), 283-307.

References

Hornbrook, M.C., M.J. Goodman, M.D. Bennett, 1991, Assessing health plan case mix in employed populations: ambulatory morbidity and prescribed drug models. In: Hornbrook M.C. (ed), Advances in health economics and health services research, 12, JAI Press inc., London, 197-232.

Howland, J., J. Stokes, S.C. Crane, A.J. Belanger, 1987, Adjusting capitation using chronic disease risk factors: A preliminary study. Health Care Financing Review, 9 (2), 15-23.

Hsia D.C., W.M. Krushat, A.B. Fagan, et al., 1988, Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. The New England Journal of Medicine, 318 (6), 352-355.

Johnson R.E., M.C. Hornbrook, G.A. Nichols, 1994, Replicating the chronic disease score (CDS) from automated pharmacy data. Journal of Clinical Epidemiology, 47 (10), 1191-1199.

Judge G.G., W.E. Griffiths, R.C. Hill, T.C. Lee, 1980, The theory and practice of econometrics. John Wiley & Sons, Inc., New York, 436-438.

Kahn, J.G., H. Luft, M.D. Smith, 1995, HIV risk adjustment: issues and proposed approaches. Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 8 (Suppl 1), S53-S66.

Kesenne, J., 1996, Health care reform and risk structure compensation, the case of Belgium. In: Proceedings of the AIM workshop on risk structure compensation, Alliance Nationale des Mutualités Chrétiennes, Brussels, 24-40.

KLOZ (Coordinating trade association of private health insurers), 1993, Circulaire KLOZ 93-069: dagverpleging (M10) (Circular: day case treatment). KLOZ, Houten.

Kronick, R., T. Dreyfus, L. Lee, Z. Zhou, 1996, Diagnostic risk adjustment for Medicaid: the disability payment system. Health Care Financing Review, 17 (2), 7-33.

Kronick, R., Z. Zhou, T. Dreyfus, 1995, Making risk adjustment work for everyone. Inquiry, 32 (1), 41-55.

Lamers, L.M., 1996, Splitsingsmodel reduceert voorspelbare winsten en verliezen (The separation of production-dependent and production-independent hospital costs reduces predictable profits and losses). Zorgverzekeraars Magazine, 2 (10), 39-40.

Lamers, L.M., 1997, Medical consumption of respondents and non-respondents to a mailed health survey. European Journal of Public Health, 7 (3), 267-271.

Lamers, L.M., B.A. van Gameren, H.C.A.M. van Vliet, 1995, De ontwikkeling van Diagnose Kosten Groepen voor de Nederlandse situatie (The development of diagnostic cost groups for the Dutch situation). instituut Beleid en Management Gezondheidszorg, Erasmus Universiteit, Rotterdam.

Lamers, L.M., R.C.J.A. van Vliet, 1996, Multiyear diagnostic information from prior hospitalizations as a risk-adjuster for capitation payments. Medical Care, 34 (6), 549-561.

Last, J.M. (Ed), 1988, A dictionary of epidemiology (second edition). Oxford university press, New York.

Lichtenstein, R., J.W. Thomas, 1987, Including a measure of health status in Medicare's HMO capitation formula: reliability issues. Medical Care, 25 (2), 100-110.

Lichtenstein, R., J.W. Thomas, J. Adams-Watson, et al., 1991, Selection bias in TEFRA at-risk HMOs. Medical Care, 29 (4), 318-331.

Light, D.W., 1992, The practice and ethics of risk-rated health insurance. JAMA, 267 (18), 2503-2508.

LSV (National Association of Specialists), 1984, Dagbehandeling (day case treatment). LSV informatiereeks.

Lubitz, J., 1987, Health status adjustment for Medicare capitation. Inquiry, 24 (4), 362-375.

Luft, H.S., 1995, Potential methods to reduce risk selection and its effects. Inquiry, 32 (1), 23-32.

Luft, H.S., R.H. Miller, 1988, Patient selection in a competitive health care system. Health Affairs, 7, 97-119.

Manton K.G., H.D. Tolley, J.C. Vertrees, 1989, Controlling risk in capitation payment, multivariate definitions of risk groups. Medical Care, 27 (3), 259-272.

Matsaganis, M., H. Glennerster, 1994, The threat of 'cream skimming' in the post-reform NHS. Journal of Health Economics, 13, 31-60.

McClure, W., 1984, On the research status of risk-adjusted capitation rates. Inquiry, 21 (3), 205-213.

Mootz, M., J. van den Berg, 1989, Indicatoren voor gezondheidstoestand in de CBS-Gezondheidsenquête (Health indicators in the CBS Health Interview Survey). Maandbericht Gezondheidsstatistiek, 8 (2), 4-10.

Nelson, L, S. Arnold, 1990, Final report on the assessment of the diagnostic cost group pilot demonstration. Health Care Financing Administration, Baltimore.

Newhouse, J.P., 1986, Rate adjusters for Medicare under capitation. Health Care Financing Review, Annual supplement, 45-55.

Newhouse, J.P., 1994, Patients at risk: health reform and risk adjustment. Health Affairs, 13, 132-146.

Newhouse, J.P., 1996, Reimbursing health plans and health providers: Efficiency in production versus selection. Journal of the Economic Literature, 34, 1236-1263.

Newhouse, J.P., W.G. Manning, E.B. Keeler, E.M. Sloss, 1989, Adjusting capitation rates using objective health measures and prior utilization. Health Care Financing Review, 10 (3), 41-54.

Newhouse, J.P., E.M. Sloss, W.G. Manning, E.B. Keeler, 1993, Risk adjustment for a children's capitation rate. Health Care Financing Review, 15 (1), 39-54.

Nooren, J.E.A.M., R.C.J.A. van Vliet, 1994, Sterfte: een verdeelkenmerk voor normuitkeringen? (Mortality: A potential risk-adjuster for capitation payments?). Tijdschrift voor Sociale Gezondheidszorg, 72 (5), 221-230.

NRV (National Council of Public Health), 1993, Dagverpleging, Advies over de ontwikkeling van dagverpleging in ziekenhuizen (Daycare, recommendations concerning the development of daycare in hospitals). NRV, Zoetermeer.

Nunnally, J.C., 1978, Psychometric theory (second edition). Mc Graw-Hill Book Company.

NZI (National Hospital Institute), 1990, Instellingen van intramurale gezondheidszorg, basisgegevens per 1-1-'90, naar provincie, naar gezondheidsregio (Institutions for inpatient care, basic information 1-1-1990). NZI, Utrecht.

NZI (National Hospital Institute), 1994, Instellingen van intramurale gezondheidszorg, basisgegevens per 1-1-'94, naar provincie, naar gezondheidsregio (Institutions for inpatient care, basic information 1-1-1994). NZI, Utrecht.

Pauly, M.V., 1984, Is cream-skimming a problem for the competitive medical market? Journal of Health Economics, 3, 87-95

PPRC (Physician Payment Review Commission), 1994a, Annual report to congress. Washington DC.

PPRC (Physician Payment Review Commission), 1994b, A Comparison of alternative approaches to risk measurement. Selected external research series, number 1, Washington DC.

Porell, F.W., W.M. Turner, 1990, Biased selection under the senior health plan prior use capitation formula. Inquiry, 27 (1), 39-50.

Riley, G., E. Rabey, J. Kasper, 1989, Biased selection and regression toward the mean in three Medicare HMO demonstrations: A survival analysis of enrollees and disrollees. Medical Care, 27 (4), 337-351.

Robinson, J.C., H.S. Luft, L.B. Gardner, E.M. Morrison, 1991, A method for risk-adjusting employer contributions to competing health insurance plans. Inquiry, 28 (2), 107-116.

Royal College of Surgeons of England, 1992, Guidelines for day case surgery (revised edition). Royal College of Surgeons of England, Commission on the Provision of Surgical Services, Londen.

Russel L.B., A. Chaudhuri, 1992, The inequality of medical expenditures for several years in a healthy, nonelderly population. Medical Care, 30 (10), 908-916.

Schauffler, H.H., J. Howland, J. Cobb, 1992, Using chronic disease risk factors to adjust Medicare capitation payments. Health Care Financing Review, 14 (1), 79-90.

Schut, F.T., 1995, Health care reform in the Netherlands: Balancing corporatism, etatism and market mechanisms. Journal of Health Politics, Policy and Law, 20 (3), 615-652.

Sheldon, T.A., P. Smith, M. Borowitz, et al., 1994, Attempt at deriving a formula for setting general practitioner fundholding budgets. British Medical Journal, 309, 1059-1064.

Shewry, S., S. Hunt, J. Ramey, J. Bertko, 1996, Risk adjustment: the missing piece of market competition. Health Affairs, 15 (1), 171-181.

SIG (Information Centre for Health Care), 1988, LMR jaarboek 1987 (National Medical Registration yearbook 1987). SIG, Utrecht.

SIG (Information Centre for Health Care), 1990, Jaarboek ziekenhuizen: kliniek en dagverpleging 1988 (Yearbook hospitals: inpatient care and daycare 1988). SIG, Utrecht.

SIG (Information Centre for Health Care), 1991a, Jaarboek ziekenhuizen: kliniek en dagverpleging 1989 (Yearbook hospitals: inpatient care and daycare 1989). SIG, Utrecht.

SIG (Information Centre for Health Care), 1991b, Jaarboek ziekenhuizen: kliniek en dagverpleging 1990 (Yearbook hospitals: inpatient care and daycare 1990). SIG, Utrecht.

SIG (Information Centre for Health Care), 1993, Jaarboek ziekenhuizen: kliniek, dagverpleging en polikliniek 1991 (Yearbook hospitals: inpatient care, daycare and outpatient care, 1991). SIG, Utrecht.

SIG (Information Centre for Health Care), 1994, Zorg in zicht (Information on the production of Dutch Hospitals). SIG, Utrecht.

SMR (Foundation Medical Registration), 1979, Classificatie van ziekten 1980, deel 1: systematisch (Classification of diseases 1980, part 1: systematic). SMR, Utrecht.

Starfield, B., J.P. Weiner, L. Mumford, D. Steinwachs, 1991, Ambulatory care groups: a categorization of diagnoses for research and management. Health Services Research, 26 (1), 53-74.

Swartz, K., 1995, Reducing risk selection requires more than risk adjustments (guest editorial). Inquiry, 32 (1), 6-10.

Swinkels, H., 1994, Trendcijfers gezondheidsenquête: medische consumptie, 1981-1993 (Trendfigures Netherlands Health Interview Survey, 1981-1993). Maandbericht Gezondheidsstatistiek, 13 (4) 4-28.

Thomas, J.W., R. Lichtenstein, 1986, Including health status in Medicare's Adjusted Average Per Capita Cost capitation formula. Medical Care, 24 (3), 259-275.

Tolley, H.D., K.G. Manton, 1984, Assessing health care costs in the elderly. Transactions of the Society of Actuaries, 36, 579-598.

Van Barneveld, E.M., L.M. Lamers, R.C.J.A. van Vliet, W.P.M.M. van de Ven, 1996, Mandatory pooling as a supplement to risk-adjusted capitation payments in a competitive health insurance market. Paper presented at the International Health Economics Association Inauguration Conference, Vancouver, Canada, May 19-23, 1996.

Van Barneveld, E.M., R.C.J.A. van Vliet, W.P.M.M. van de Ven, 1997, Risk-adjusted capitation payments for catastrophic risks based on multi-year prior costs. Health Policy, 39, 123-135.

Van de Ven, W.P.M.M., F.T. Schut, 1994, Should catastrophic risks be included in a regulated competitive health insurance market? Social Science & Medicine, 39 (10), 1459-1472.

Van de Ven, W.P.M.M., R.C.J.A. van Vliet, 1992, How can we prevent cream skimming in a competitive health insurance market? The great challenge of the 90's. In: French, T. and P. Zweifel (eds), Health Economics Worldwide, Kluwer Academic Publishers, Amsterdam.

Van Vliet, R.C.J.A., 1992, Predictability of individual health care expenditures. Journal of Risk and Insurance, 59 (3), 443-461.

Van Vliet, R.C.J.A., W.P.M.M. van de Ven, 1992, Towards a capitation formula for competing health insurers. An empirical analysis. Social Science & Medicine, 34 (9), 1035-1048.

Van Vliet, R.C.J.A., W.P.M.M. van de Ven, 1993, Capitation Payments based on prior hospitalizations. Health Economics, 2, 177-188.

Von Korff, M., E.H. Wagner, K. Saunders, 1992, A chronic disease score from automated pharmacy data. Journal of Clinical Epidemiology, 45 (2), 197-203.

VWS (Ministry of Health, Welfare and Sports), 1995, De prijs die zorg verdient, prijsbeleid en financiaal beleid zorgsector (Policy on prices and financing of the health care sector). TK 24478 (1), SDU, The Hague.

Weiner, J.P., A. Dobson, S.L. Maxwell, et al., 1996, Risk-adjusted Medicare capitation rates using ambulatory and inpatient diagnoses. Health Care Financing Review, 17 (3), 77-99.

Weiner, J.P., B.H. Starfield, D.M. Steinwachs, L.M. Mumford, 1991, Development and application of a population-oriented measure of ambulatory care case-mix. Medical Care, 29 (5), 452-472.

Wouters, A.V., 1991, Disaggregated annual health services expenditures and role as predictors. Health Services Research, 26 (2), 247-272.

WVC (Ministry of Welfare, Health and Cultural Affairs), 1988, Changing health care in the Netherlands. WVC, Rijswijk.

ZFR (Sickness Fund Council), 1985, Voorlopige indicatielijsten verrichting dagverpleging, ontleend aan circulaire, nr 13774-146/85 (Interim list procedures day case treatment). ZFR, Amstelveen.

ZFR (Sickness Fund Council), 1995a, Rapportage projectgroep experiment dubbele budgettering (Report on the separation of production-dependent and production-independent hospital costs). ZFR, Amstelveen.

ZFR (Sickness Fund Council), 1995b, Besluit budgettering verstrekkingen ZFW 1996 (Regulations for the allocation of budgets for the social health insurance sector 1996). ZFR, Amstelveen.

Samenvatting

Normuitkeringen voor ziekenfondsen gebaseerd op diagnose-informatie van ziekenhuisopnamen in het verleden

In Nederland zijn eind jaren tachtig veranderingen in de structuur en financiering van de gezondheidszorg in gang gezet. Het doel van deze hervormingen was de kosten van de gezondheidszorg te beheersen en te komen tot een grotere doelmatigheid. Daarbij moest het sociale karakter van de gezondheidszorg behouden blijven. Een belangrijk onderdeel van deze hervormingen was het creëren van een concurrerende verzekeringsmarkt, waarbij zorgverzekeraars geprikkeld worden om verzekerden aan te trekken door goede kwaliteit zorg te bieden en een efficiënte werkwijze te hanteren.

Op een concurrerende verzekeringsmarkt zullen de premies het gezondheidsrisico van de verzekerden weergeven. Dit kan betekenen dat de premie voor een tachtigjarige met een aantal chronische aandoeningen tientallen malen hoger is dan de premie voor een gezonde twintiger. Vanuit maatschappelijk oogpunt bezien is dit een onwenselijke situatie. Om de toegang tot de gezondheidszorg te garanderen op een concurrerende verzekeringsmarkt is een stelsel van normuitkeringen ingevoerd.

Ziekenfondsverzekerden betalen een grotendeels inkomensafhankelijke premie voor hun ziektekostenverzekering. Deze premies worden verzameld in de zogenaamde 'Algemene Kas van de Ziekenfondswet'. Tot 1993 kregen de ziekenfondsen alle kosten vergoed, die door hun verzekerden werden gemaakt. Sinds 1993 ontvangen de ziekenfondsen uit de Algemene Kas normuitkeringen ter bekostiging van de verstrekkingen die vallen onder de Ziekenfondswet. De hoogte van de normuitkering is afhankelijk van de leeftijd en het geslacht van de verzekerde, de regio waarin de verzekerde woont en of hij/zij al dan niet arbeidsongeschikt is. Naast

normuitkeringen ontvangen de ziekenfondsen ook nominale premies. Deze nominale premies betalen de verzekerden rechtstreeks aan het ziekenfonds van hun keuze. De hoogte van de nominale premie kan per ziekenfonds verschillen, maar is voor alle verzekerden van één ziekenfonds gelijk. De hoogte van de nominale premie weerspiegelt het succes van het ziekenfonds met betrekking tot kostenbeheersing en doelmatigheid.

Met het stelsel van normuitkeringen wordt beoogd zo goed mogelijk de aanvaardbare kosten te voorspellen van verschillen in medische consumptie die voortvloeien uit behoeften aan zorg. Het is van groot belang dat de normuitkeringen zoveel mogelijk aansluiten op het gezondheidsrisico-profiel van de verzekerden. Indien dit niet het geval is kunnen individuele ziekenfondsen met relatief veel slechte risico's in financiële problemen komen. Daarnaast zal dit ertoe leiden dat chronisch zieken, vanwege de voorspelbare verliezen die zij genereren, vanuit financieel oogpunt bezien bij elk ziekenfonds ongewenste klanten zijn. Het stelsel van normuitkeringen moet ziekenfondsen stimuleren doelmatiger te gaan werken en niet aanzetten tot het toepassen van gunstige-risicoselectie.

In 1993 was een ziekenfonds financieel verantwoordelijk voor circa 3% van het verschil tussen zijn werkelijke uitgaven en de genormeerde uitgaven; 97% van de verschillen werden achteraf verevend en nagecalculeerd. In 1997 is de financiële verantwoordelijkheid van ziekenfondsen toegenomen. Na verevening en nacalculatie dragen ziekenfondsen een financieel risico van 5% voor de door hen niet te beïnvloeden vaste ziekenhuiskosten en van circa 52,5% voor de variabele, productiegebonden ziekenhuiskosten en de kosten van overige verstrekkingen. Daarnaast geldt voor de variabele ziekenhuiskosten en kosten van overige verstrekkingen een overschadevergoeding van 90% voor kosten van individuele verzekerden die boven de 4.500 gulden uitkomen. In totaal betekent dit dat een ziekenfonds in 1997 een financiële verantwoordelijkheid heeft van circa 25% van het verschil tussen zijn werkelijke uitgaven en de genormeerde uitgaven. Het ligt in de bedoeling van het kabinet Kok om de financiële verantwoordelijkheid van ziekenfondsen te vergroten tot ongeveer 55% in 1998. Een noodzakelijke voorwaarde om dit doel op een maatschappelijk verantwoorde wijze te bereiken, is een verdere verbetering van het stelsel

van normuitkeringen. Daarnaast dienen de sturingsmogelijkheden van de ziekenfondsen te worden vergroot. Dit proefschrift richt zich op het verbeteren van het stelsel van normuitkeringen door het verfijnen van de verdeelformule, waarop de normuitkeringen zijn gebaseerd.

Potentiële verdeelkenmerken voor normuitkeringen

De verdeelkenmerken op basis waarvan de normuitkeringen worden bepaald, dienen zo te worden gekozen, dat zij de verzekerden indelen in wat betreft zorgbehoefte homogene groepen. De verdeelkenmerken moeten voorspellende waarde hebben voor toekomstige ziektekosten. Daarnaast moeten ze betrouwbaar gemeten kunnen worden, niet manipuleerbaar zijn, de juiste prikkels geven en beschikbaar zijn. Geen enkel verdeelkenmerk zal volledig aan al deze voorwaarden voldoen. De thans als verdeelkenmerk gebruikte demografische factoren zijn beschikbaar, niet te manipuleren, betrouwbaar te meten, maar het zijn slechte proxy's van behoefte aan zorg en daarmee geen goede voorspellers van toekomstige ziektekosten op individueel niveau. Toevoeging van meer direct aan zorgbehoefte gerelateerde verdeelkenmerken kan de werking van het huidige demografische verdeelmodel verbeteren. Als veelbelovende verdeelkenmerken komen uit Amerikaanse studies naar voren het gebruik van zorgvoorzieningen in het verleden, al dan niet in combinatie met diagnose-informatie, ervaren gezondheid en functionele gezondheidstoestand.

Zorggebruik in het verleden is de beste voorspeller van toekomstige ziektekosten. Echter, toepassing van dergelijke informatie als basis voor een verdeelkenmerk heeft een aantal nadelen. Het gevaar bestaat dat inefficiënte ziekenfondsen worden beloond en het kan aanmoedigen tot meer gebruik dan noodzakelijk is. De mate waarin een verdeelkenmerk gebaseerd op zorggebruik in het verleden deze perverse prikkels oplevert, zal afhangen van de marginale kosten voor onnodig en/of inefficiënt zorggebruik en de verwachte marginale opbrengsten in termen van een hogere toekomstige normuitkering. Deze nadelen kunnen gedeeltelijk worden ondervangen door niet het zorggebruik op zich, maar de diagnose-informatie van

zorggebruik in het verleden als verdeelkenmerk te hanteren. In diverse studies zijn verdeelkenmerken ontwikkeld gebaseerd op diagnose-informatie van ziekenhuisopnamen, gebaseerd op diagnoses van niet-klinisch zorggebruik in ziekenhuizen en gebaseerd op chronisch aandoeningen afgeleid uit medicijngebruik. Deze verdeelkenmerken hebben voorspellende waarde voor toekomstige ziektekosten. Het toepassen van dergelijke verdeelkenmerken kan ook onnodig gebruik van de betreffende voorzieningen stimuleren om in een volgend jaar een hogere normuitkering te verkrijgen. Daarnaast kan zogenaamde diagnose-'inflatie' optreden. Dit houdt in dat veranderingen in het coderen van diagnoses optreden. Verschuivingen zullen optreden in de richting van diagnosecategorieën die een hogere normuitkering opleveren.

Naast het gebruik van zorgvoorzieningen in het verleden zijn ervaren gezondheid, functionele gezondheidstoestand en chronische aandoeningen goede voorspellers van toekomstige ziektekosten. Diverse onderzoeken laten zien dat toevoeging van deze factoren, apart of in combinatie met elkaar en met zorggebruik in het verleden, de voorspelkracht van een verdeelmodel met alleen demografische variabelen, substantieel verbetert. Gegevens over ervaren gezondheid, functionele gezondheidstoestand en (zelf-gerapporteerde) chronische aandoeningen worden veelal verkregen door middel van enquêtes. Over het algemeen komt enquête-informatie niet voor in de administratie van ziekenfondsen. Deze informatie lijkt als verdeelkenmerk voor normuitkeringen in de Nederlandse context dan ook niet geschikt.

Dit proefschrift richt zich op het toepassen van diagnose-informatie van ziekenhuisopnamen in het verleden als verdeelkenmerk voor normuitkeringen. Dit verdeelkenmerk vormt de beste combinatie van voorspellende waarde voor toekomstige ziektekosten, beschikbaarheid en juiste prikkels, eigenschappen die verdeelkenmerken idealiter bezitten. Bij veel ziekenfondsen werden bij aanvang van het onderzoek de diagnoses van ziekenhuisopnamen routinematig vastgelegd in de administratie. Het Diagnose Kosten Groepen (DKG) model wordt gebruikt om deze diagnose-informatie in een verdeelmodel voor normuitkeringen op te nemen.

Het Diagnose Kosten Groepen model

Het DKG model is ontwikkeld in de Verenigde Staten met gegevens, welke hoofdzakelijk betrekking hebben op bejaarden. Informatie over diagnoses die zijn gesteld bij ziekenhuisopnamen in het verleden, worden geclusterd tot DKGs. Bij deze clustering is zowel van medische als economische criteria gebruik gemaakt. In eerste instantie zijn opnamediagnoses bij ziekenhuisopname (in de vorm van 3-cijferige ICD-codes) gegroepeerd tot 78 klinisch gezien homogene subgroepen. Vervolgens zijn per subgroep de gemiddelde 'vervolgkosten' -dat wil zeggen de ziektekosten in het direct volgende jaar van patiënten met de betreffende diagnoses- bepaald. Op grond van overeenkomsten in gemiddelde vervolgkosten zijn de 78 subgroepen samengevoegd tot negen DKGs. Personen die niet zijn opgenomen in een ziekenhuis worden in DKG 0 geplaatst. De voorspelkracht van een verdeelmodel waarin diagnose-informatie in de vorm van DKGs is opgenomen naast demografische variabelen is aanzienlijk beter dan de voorspelkracht van een model met alleen demografische variabelen. In dit proefschrift is het oorspronkelijke Amerikaanse DKG model toegepast op gegevens van ziekenfondsverzekerden van alle leeftijden. Ook voor Nederlandse ziekenfondsverzekerden verbetert de voorspelkracht van het demografische verdeelmodel aanzienlijk na uitbreiding met DKGs.

In dit proefschrift is de ontwikkeling beschreven van DKGs voor de Nederlandse situatie. Voor de empirische analyses is gebruik gemaakt van een cohortbestand met verzekerings-, opname- en schadegegevens over de jaren 1988-1994 van ruim 50.000 ziekenfondsverzekerden van alle leeftijden. Allereerst is nagegaan of de clustering van diagnoses op basis van 'Nederlandse' vervolgkosten tot negen DKGs een DKG-indeling oplevert met een grotere voorspelkracht dan de oorspronkelijke Amerikaanse indeling. Daarbij zijn drie typen vervolgkosten onderscheiden: de vervolgkosten in het jaar na opname, de voor leeftijd en geslacht gecorrigeerde vervolgkosten in het jaar na opname en de vervolgkosten in het tweede en derde jaar na opname.

De verschillen in voorspelkracht tussen de DKG-indelingen gebaseerd op verschillende Nederlandse vervolgkosten onderling en met de Amerikaanse indeling zijn gering. Op inhoudelijke gronden is gekozen voor een DKG-indeling op basis van vervolgkosten in het tweede en derde jaar na opname. Door het eerste jaar na opname over te slaan, worden immers vervolgkosten voor eenmalige gezondheidsproblemen, waarvoor in sommige gevallen toch een kortdurende vervolgbehandeling of een aantal controles nodig zijn, buiten beschouwing gelaten. Op deze wijze vindt de indeling in DKGs plaats op basis van kosten die, naar verwachting, voortkomen uit chronische gezondheidsproblemen. Het aantal DKGs kan van negen nog worden teruggebracht tot vijf zonder dat dit tot een noemenswaardig verlies in voorspelkracht leidt.

Een volgende belangrijke onderzoeksvraag was: wat is in het kader van normuitkeringen mede gebaseerd op DKGs, het effect van het plaatsen van diagnoses geïndiceerd voor dagopname in DKG 0 bij de groep personen zonder opname? Met het oog op bevordering van doelmatigheid en kostenbeheersing is het wenselijk substitutie van ziekenhuisopname door dagbehandeling te stimuleren. Bij de indeling in DKGs dienen daarom personen met een ziekenhuisopname met een ICD-code waarvoor een dagopname geïndiceerd is, in DKG 0 te worden geplaatst, dat wil zeggen dat opnamen met die diagnoses niet leiden tot een hogere normuitkering.

Aangezien in Nederland geen algemeen geaccepteerde lijsten voorhanden zijn van medisch gezien verantwoord in dagbehandeling uit te voeren verrichtingen, is op basis van een empirische analyse van alle ziekenhuisopnamen in Nederland een aantal lijsten samengesteld met diagnoses waarvoor een dagopname geïndiceerd lijkt te zijn. In deze analyse is nagegaan wat in de praktijk is gerealiseerd in 1993. Hiervoor is gebruik gemaakt van een databestand dat alle ziekenhuisopnamen, zowel dagopnamen als klinische opnamen, in 1993 in Nederland bevat. In 1993 betrof ruim een kwart van alle opnamen een dagopname. De voorspelkracht van DKG modellen vermindert enigszins wanneer diagnoses geïndiceerd voor dagopname in DKG 0 worden geplaatst bij de personen zonder ziekenhuisopname. Geconcludeerd is dat het niet tot een noemenswaardig verlies aan voorspelkracht leidt wanneer opnamen voor

diagnoses die in meer dan 35% van de gevallen in dagopname plaatsvinden (ongeveer 36% van alle klinische opnamen in het cohortbestand) in DKG 0 worden geplaatst bij de personen zonder ziekenhuisopname.

Om een valide verdeelkenmerk voor normuitkeringen te zijn, moeten DKGs niet alleen toekomstige ziektekosten kunnen voorspellen, maar ook de gezondheidstoestand van de verzekerden weergeven. Indien DKGs de gezondheidstoestand van de verzekerden meten, moeten zij samenhang vertonen met gezondheidsindicatoren. Uit een analyse, waarvoor gebruik is gemaakt van een gezondheidsenquête gehouden onder 15.000 personen uit het totale cohortbestand, blijkt dit inderdaad het geval te zijn. De relatie tussen DKGs en de gezondheidsindicatoren is zoals verwacht: hoe hoger de DKG, des te slechter de gezondheidstoestand. De sterkste samenhang vertonen de DKGs met meer algemene gezondheidsindicatoren zoals ervaren gezondheid en het aantal zelfgerapporteerde langdurige aandoeningen.

Meerjarige DKGs

In de Amerikaanse onderzoeken zijn ziekenhuisopnamen in een periode van één jaar gebruikt om te komen tot een toewijzing van verzekerden aan DKGs. Verzekerden met chronische aandoeningen en daarmee samenhangende structureel hoge ziektekosten worden echter niet elk jaar in een ziekenhuis opgenomen. De kans om deze verzekerden toch in een DKG in te delen, wordt groter naarmate de DKGs worden gebaseerd op een langere periode. Met diagnose-informatie van ziekenhuisopnamen in de afgelopen drie jaar (het driejarig DKG model) blijken toekomstige ziektekosten beter te kunnen worden voorspeld dan met diagnose-informatie van het afgelopen jaar (het éénjarige DKG model). Dit blijkt onder meer uit de resultaten voor een aantal risicogroepen gedefinieerd op basis van medische consumptie en kosten in 1988. Het verschil tussen feitelijke en voorspelde kosten in 1992 voor personen met een ziekenhuisopname in de hoogste DKGs in 1988 vermindert van 209% (demografisch model) tot 125% indien de kosten worden voorspeld met het éénjarig DKG model en tot 74% wanneer het driejarig DKG model wordt

gebruikt om de kosten te voorspellen. Voor de 5,5% van de verzekerden met de hoogste kosten in 1988 liggen de feitelijke kosten in 1992 88% boven de voorspelde kosten van het demografische model. Het verschil tussen feitelijke en voorspelde kosten reduceert tot 59% en 42% wanneer de kosten worden voorspeld met het éénjarig- respectievelijk driejarig DKG model. Met name voor verzekerden met een hoge medische consumptie in 1988 kan de voorspelkracht van de DKG modellen nog verder worden verbeterd indien tevens informatie over hoge kosten in het verleden in het model wordt opgenomen.
Uitgaande van het vigerende demografische verdeelmodel blijkt op basis van diagnose-informatie van ziekenhuisopnamen in de afgelopen drie jaar 15% van de verzekerden een groep zogenaamde slechte risico's te vormen. Het voorspelbare verlies voor deze verzekerden bedraagt gemiddeld bijna 1900 gulden per verzekerde. Binnen deze groep is het voorspelbare verlies voor degenen met een ziekenhuisopname in het direct voorgaande jaar ruim 3000 gulden.

Tot slot is de voorspelkracht van DKG modellen geëvalueerd, waarbij gebruik werd gemaakt van gegevens van de gezondheidsenquête. Uit deze evaluatie blijkt dat een set van acht variabelen uit de gezondheidsenquête de voorspelkracht van een driejarig DKG model nog kan verbeteren. Binnen de groep van personen met een ziekenhuisopname in het direct voorgaande jaar kan op basis van de geselecteerde enquête-informatie nog een nader onderscheid worden gemaakt in goede en slechte risico's, dat wil zeggen in verzekerden met voorspelbare winsten en voorspelbare verliezen op basis van de enquête-informatie. Dit geldt zowel uitgaande van een demografisch verdeelmodel als voor de DKG modellen. Van de personen, die op basis van de enquête-informatie als slecht risico zijn aan te wijzen, is het overgrote deel niet in het ziekenhuis opgenomen. Dit lijkt een groep te zijn met een chronische behoefte aan zorg en daarmee samenhangende voorspelbare hoge ziektekosten. Deze personen worden niet in een DKG ingedeeld en hun kosten worden op basis van alleen leeftijd, geslacht, regio en arbeidsongeschiktheid slecht voorspeld. Zij ervaren hun gezondheid als matig tot slecht, ondervinden mogelijk beperkingen in hun lichamelijk functioneren, hebben frequent contact met hun huisarts, gebruiken veel geneesmiddelen en maken gebruik van wijkverpleging. Hun ziektekosten

zullen hoger zijn dan die van gezonde leeftijdsgenoten. Om ook rekening te houden met de voorspelbaar hoge kosten van deze verzekerden kan worden gedacht aan een verdeelkenmerk gebaseerd op diagnose-informatie van niet-klinisch zorggebruik.

Een normuitkeringensysteem dat mede is gebaseerd op diagnose-informatie van ziekenhuisopnamen is wat betreft uitvoerbaarheid complexer dan het huidige systeem gebaseerd op een demografisch verdeelmodel. Invoering van een DKG model zal de voorspelkracht voor toekomstige ziektekosten aanzienlijk verbeteren. Daar staat tegenover dat monitoring zal moeten plaatsvinden om diagnose-inflatie en onnodige ziekenhuisopnamen te voorkomen.
De voorspelkracht van het driejarige DKG model is groter dan van het éénjarige DKG model. Echter, wat betreft de uitvoerbaarheid zal een driejarig DKG model meer problemen met zich meebrengen. Zo moeten ziekenfondsen over drie jaar diagnose-informatie aanleveren aan de Algemene Kas. Ziekenfondsverzekerden kunnen in principe elk jaar van ziekenfonds wisselen. Dit kan de praktische uitvoerbaarheid van normuitkeringen gebaseerd op een driejarig DKG model bemoeilijken. Het is derhalve aan te bevelen om met de invoering van een éénjarig DKG model te beginnen. Indien wordt gevreesd voor een toename van ziekenhuisopnamen voor diagnoses die in principe in dagopname kunnen worden behandeld, kan worden overwogen om alleen opnamen in de hogere DKGs, bijvoorbeeld vanaf DKG 3, te laten leiden tot een hogere normuitkering.

Bij aanvang van dit onderzoek in 1992 kon ieder ziekenfonds in principe beschikken over diagnose-informatie van ziekenhuisopnamen op individueel niveau. Anno 1997 is dit niet langer het geval. Na bepaling van de gewichten voor DKGs is het voor de uitvoerbaarheid van een systeem van normuitkeringen echter niet noodzakelijk dat een ziekenfonds beschikt over diagnose-informatie op individueel niveau. Een normuitkeringensysteem mede gebaseerd op DKGs is uitvoerbaar indien de Algemene Kas voor iedere verzekeraar kan beschikken over het aantal verzekerden per DKG.

Verzekeraars hebben op grond van de voor hen relatief eenvoudig toegankelijke informatie over ziekenhuisopnamen en (hoge) kosten in het verleden een voorsprong op de Algemene Kas bij het huidige verdeelmodel. Zij kunnen subgroepen met voorspelbare winsten en verliezen in principe onderscheiden aan de hand van ziektekosten in het verleden, al dan niet in combinatie met diagnose-informatie. In 1997 geldt een overschadevergoeding van 90% voor de variabele ziekenhuiskosten plus de kosten van overige verstrekkingen boven de 4.500 gulden. Deze regeling vermindert de prikkels tot gunstige-risicoselectie aanzienlijk, maar neemt ook de prikkels tot kostenbeheersing en doelmatigheid weg voor verzekerden met hoge kosten. Indien het systeem van normuitkeringen mede gebaseerd wordt op DKGs worden de prikkels tot gunstige-risicoselectie eveneens verminderd, maar blijven de prikkels tot doelmatigheid ook voor verzekerden met hoge kosten bestaan. Hoewel ook het DKG model niet perfect is, kan het een substantieel deel van individuele verschillen in ziektekosten voorspellen. Het gebruik van diagnose-informatie van ziekenhuisopnamen in het verleden biedt dan ook aanknopingspunten om het verdeelmodel verder te verfijnen en daarmee het stelsel van normuitkeringen te verbeteren.

APPENDIX A

Medical consumption of respondents and non-respondents to a mailed health survey [22]

[22] This paper has been published in the European Journal of Public Health (Lamers, 1997).

ABSTRACT

Non-response bias can distort the results of health surveys. The occurrence of selective non-response can be assessed when data are available for both respondents and non-respondents. The objective of this study was to compare the medical consumption of respondents and non-respondents to a mailed health survey. A mailed health survey was conducted among approximately 13,500 adults and among parents of approximately 1,500 children aged 5-15 years. The net response rate was 70.4%. A panel data set that could be matched with the health survey data was available for all eligible persons. This data set comprises administrative information on hospitalizations, annual health care expenditures and demographic variables. The results of this study show that response was associated with age, sex, degree of urbanization and type of insurance. After correcting for differences in demographic variables, respondents and non-respondents differ in the utilization of several types of care. Relatively more users than non-users responded. Response was not associated with the utilization of care related to severe conditions such as in-patient hospital care. The conclusion from this study is that when a mailed health survey is used to measure medical consumption, non-response bias will result in a small overestimation of utilization.

INTRODUCTION

Non-response or selection bias refers to situations in which a sample differs in some systematic way from the population from which it was drawn, e.g. data are missing in a non-random fashion. Selection bias can distort the results of health surveys. The occurrence of selective non-response can be assessed when (administrative) data are available for both respondents and non-respondents. Comparisons of respondents and non-respondents are often restricted to demographic variables such as age, sex, place of residence and degree of urbanization.

Most studies consider the relation between age and survey participation. Elderly persons are more likely than younger persons to refuse or fail to participate in surveys[1-4] or to answer specific questions.[5] Another study showed that response rate increases with age, up to 50 years.[6] Regional differences in response rates are common. Non-response is also associated with the degree of urbanization. Lower response rates occur in urban areas.[1,4] The relationship between sex and response rate is not always consistent. Higher levels of non-response among males are reported.[4] In a community survey of health and social status in persons aged 65 years and older the decision to participate was gender neutral.[3] The same study also showed that health and living alone or with others favouring participation were positively associated with participation. The relation between non-response and family size is not consistent. In one study no relation between response and the number of adults in the household was found,[7] while in another a positive association between the number of family members and the response rate occurred.[1]

Health may be an important determinant of response. In a population-based cardiovascular disease study, risk factors distinguish respondents from non-respondents in only some cases.[8] Respondents and non-respondents to a mailed health survey among participants in a longitudinal study of physical activity, physical fitness and health were equally healthy at entry to the study. Those who had family members with chronic conditions and who had positive health behaviors were more likely to respond to the health survey.[9] In a Danish study on nutrition and health in the elderly the self-perceived health status was a determining factor in the decision to participate. Participation was less likely the lower the self-judged health. More non-participants had been hospitalized in the year before the study.[10] A negative association between hospital and nursing home utilization and the response rate was also reported for a survey of medicine use among elderly.[11] Persons with ambulatory physician visits were significantly more likely to respond.

In this article the medical consumption of respondents and non-respondents to a mailed health survey is compared. First, differences between these groups with regard to age, sex and degree of urbanization are examined. Second, while correcting for differences in these demographic variables, respondents and non-respondents are compared with regard to the utilization of several types of care.

METHODS

Data

In February 1993 a first mailing, containing the questionnaire with the cover letter, was sent. After one week everyone received a postcard reminder, which served both as a 'thank you' for those who had already responded and as a friendly reminder for those who had not. Three weeks after this postcard reminder a letter with a new questionnaire was sent to the non-respondents only. Two weeks after this, the non-respondents received a final mailing. The design of this procedure was guided by Dillman's[12] recommendations.

The main purpose of the survey was to gather information on health status and (additional) medical consumption to be used for a study on risk-adjusted capitation payments to health insurers, an important issue in Dutch health care reforms. Approximately 15,000 people received the health survey. Among them were 13,472 adults between 15 and 90 years old and parents of 1,509 children aged 5-14 years. The parents were asked to complete the questionnaire for their child. All eligible persons were enrollees from one social health care insurer, a so-called sickness fund, in the Western part of The Netherlands. In The Netherlands the sickness funds provide compulsory health insurance coverage for the approximately 60% of the population in the lowest income brackets. A panel data set containing administrative information that could be matched with the health survey data was available for all eligible persons. For both respondents and non-respondents the panel data set comprises administrative information on hospitalizations (when applicable), annual health care expenditures and demographic variables such as age, sex, postcode and type of insurance for five years, from 1988 to 1992.

The survey data were compared with these administrative data for date of birth and sex to make sure that the eligible person completed the questionnaire and not someone else in the household. For the adults 413 questionnaires appeared to be completed by another person; 64 parents completed the questionnaire for another child than the one asked for. Questionnaires completed by other than the eligible person are considered as non-responses, resulting in a net response rate of 70.0%

for adults and of 75.4% for children. The net response rate for the total sample is 70.4%.

The questionnaire contained questions about health and medical consumption. The health indicator questions refer to perceived health status, chronic conditions, functional disabilities and vague psychosomatic complaints. The questionnaire also contained questions about date of birth, sex, the respondent's country of birth and that of his/her parents, education, marital status and the number of persons in the household.

The questions about medical consumption refer to consultation with a GP during the previous two months and consultation with a medical specialist, physiotherapist, speech therapist and dentist during the previous 12 months. The respondents were also asked how many times they had been hospitalized during the previous 12 months and whether they had used prescribed drugs during the previous 14 days.

Analysis

Administrative data on health care expenditures in 1992 and the demographic variables age, sex, degree of urbanization and type of insurance were used for comparison between respondents and non-respondents. To assess the relation between non-response and demographic variables a logistic regression analysis was applied.[13] In this analysis age*sex, degree of urbanization and type of insurance were the independent variables and response the dependent variable. Fifteen dummy variables for age*sex (seven 10-year and one 15-year age groups for each sex minus 1) were included in the logistic regression model as well as four dummy variables for degree of urbanization and three dummy variables for type of insurance. For degree of urbanization the address density of the surrounding area was used.[14] Type of insurance is the compulsory cause for enrolment with the sickness fund, for example wage earners with a salary below a certain threshold, recipients of disability or unemployment benefits and elderly people with low incomes. This variable was included in the model because of its importance for the study on risk-adjusted capitation payments for which the survey data were primarily gathered. Every member of a family with an income below a certain threshold is a policy holder. Thus, within one family more than one policy holder can occur. In the Dutch social health insurance system deductibles and co-payments were virtually absent in 1993, the year that the survey was held.

To assess the relation between non-response and medical consumption another 10 logistic regression analyses were applied. In these analyses age*sex, degree of urbanization, type of insurance and a dummy variable for utilization of a certain

type of care in 1992 were the independent variables and response the dependent variable. Utilization in 1992 was derived from the costs for the relevant types of care in the administrative data. Ten types of care could be distinguished. For each type of care the relation of utilization with response was examined separately.

RESULTS

The distributions of age, sex, degree of urbanization and type of insurance for both the respondents and the non-respondents are shown in table A.1. Young persons of 15-24 years old and elderly persons aged 75 years and older, men, persons living in highly urbanized regions and disabled and other policy holders were found relatively more often among non-respondents.

Table A.1 Distribution of age, sex, degree of urbanization and type of insurance for respondents and non-respondents

	Respondents (%)	Non-respondents (%)	Total (%)
Age (years)			
5-14	10.7	8.5	10.0
15-24	12.5	15.1	13.3
25-34	22.0	24.1	22.6
35-44	16.1	15.1	15.8
45-54	13.2	12.4	13.0
55-64	11.5	9.0	10.8
65-74	8.6	7.8	8.4
75-90	5.4	8.0	6.1
Sex			
Male	41.2	50.8	44.1
Female	58.8	49.2	55.9
Degree of urbanization			
Very strongly urban/big city	11.7	17.3	13.4
Strongly urban	24.4	26.2	24.9
Moderately urban	36.2	32.9	35.2
Little urban	17.0	15.3	16.5
Rural	10.7	8.3	10.0
Type of insurance			
Employed policy holder	46.8	46.0	46.6
Disabled policy holder	5.3	6.8	5.7
Other policy holder	20.7	24.5	21.9
Family member	27.2	22.7	25.9

The demographic variables were used in a logistic regression analysis with response as the dependent variable. The largest groups were used as the reference groups, i.e. women aged 25-34 years, persons in moderately urbanized (category 3) regions and employed policy holders. Table A.2 shows the results of the logistic regression analysis. These results are adjusted for all variables in the table.

Table A.2 Odds ratios and 95% confidence intervals for participation in survey

	Odds ratio	95% confidence interval
Age, in women (years)		
5-14	0.98	0.76-1.19
15-24	0.95	0.77-1.12
25-34	1	
35-44	1.18	1.02-1.35
45-54	1.19	1.02-1.37
55-64	1.45	1.25-1.64
65-74	1.44	1.21-1.68
75-90	0.83	0.59-1.06
Age, in men		
5-14	1.11	0.89-1.33
15-24	0.44	0.27-0.61
25-34	0.53	0.38-0.68
35-44	0.62	0.44-0.79
45-54	0.64	0.45-0.83
55-64	1.21	0.99-1.42
65-74	1.54	1.27-1.80
75-90	1.15	0.85-1.45
Degree of urbanization		
Very strongly urban / big city	0.63	0.52-0.75
Strongly urban	0.84	0.75-0.94
Moderately urban	1	
Little urban	1.02	0.91-1.13
Rural	1.17	1.04-1.31
Type of insurance		
Employed policy holder	1	
Disabled policy holder	0.58	0.42-0.74
Other policy holder	0.56	0.41-0.71
Family member	0.89	0.77-1.01

Table A.2 shows that selectivity in response occurred. Relatively more respondents were women, aged 35-74 years in comparison to the reference group. Men in the age group of 15-54 years were relatively less often among the

respondents, while relatively more men, aged 65-74 years participated in the survey. Persons living in the (highly) urbanized regions were relatively less often found among respondents than persons living in moderately urbanized regions, while persons living in rural regions were relatively more often respondents. For type of insurance table A.2 shows that disabled and other policy holders belonged relatively more often to the non-respondents than the employed policy holders. Table A.3 shows the proportion of persons with costs in 1992 for several types of care for both respondents and non-respondents. For most types of care the percentages of users are higher among respondents. The odds ratios for utilization from the logistic regression analyses with utilization in 1992 and the demographic factors as independent variables and response as the dependent variable are also presented in table A.3. The groups without utilization were the reference groups. The odds ratios were corrected for differences in all four demographic variables. Relatively more users than non-users of out-patient specialist care, prescribed drugs, obstetrics, paramedical services (mostly physiotherapy), maternity care, technical aids and dental care were respondents. For the other types of care the odds ratios did not differ significantly from 1.

Table A.3 Distribution of utilization in 1992 for respondents and non-respondents and odds ratios and 95% confidence intervals for participation in survey

	Respondents (%)	Non-respondents (%)	Odds ratio[b]	95% confidence interval[b]
Utilization in 1992 per type of care				
In-patient care	6.4	6.6	0.98	0.83-1.13
Out-patient specialist care	50.7	44.0	1.29	1.21-1.36
Prescribed drugs	80.2	75.9	1.21	1.12-1.30
Obstetrics	1.4	0.6	2.03	1.59-2.46
Maternity care	1.5	0.6	2.11	1.69-2.52
Paramedical services	18.9	15.4	1.27	1.17-1.37
Dental care [a]	58.9	50.8	1.41	1.33-1.49
Technical aids	6.9	5.5	1.34	1.18-1.50
Transport by ambulance	2.0	2.7	0.81	0.57-1.05
Other transport	1.5	1.6	1.07	0.76-1.36

[a] Approximately one-quarter of the costs for dental care in 1992 are only available in the sickness fund administration on an aggregated level and can not be matched with individual records. This affects the percentages of dental care users of both respondents and non-respondents. The percentages presented are underestimations of dental care utilization.

[b] Odds ratios and 95% confidence intervals corrected for age*sex, degree of urbanization and type of insurance.

For the logistic regression analyses utilization was defined as whether or not persons had costs for the relevant types of care. In table A.4 the mean costs per type of care are shown for respondents and non-respondents. T-tests were performed to assess the differences in the mean costs between the two groups. Respondents had significantly higher mean costs than non-respondents for obstetrics, maternity care, paramedical services and dental care. For the other types of care no differences in the mean costs occurred. This comparison was not corrected for differences in demographic variables. However, using direct standardization or post-stratification[15] with regard to age, sex, degree of urbanization and type of insurance to adjust for non-response bias related to these variables hardly affected the results.

Table A.4 Mean costs[a] in 1992 and standard deviations by type of care for respondents and non-respondents

	Respondents		Non-respondents		Test results[b]	
	Mean	SD	Mean	SD	*t*-value	*p*-value
Type of care						
In-patient care	620	3,973	628	3,996	- 0.10	0.92
Out-patient specialist care	293	1,790	239	1,453	1.69	0.09
Prescribed drugs	377	1,046	388	938	- 0.55	0.58
Obstetrics	8	75	3	45	4.92	0.00
Maternity care	57	1,320	20	330	2.55	0.01
Paramedical services	121	380	105	404	2.12	0.03
Dental care	75	168	64	174	3.33	0.00
Technical aids	39	315	32	283	1.30	0.19
Transport by ambulance	16	153	20	141	- 1.53	0.13
Other transport	14	229	12	195	0.28	0.79

[a] Costs are in Dutch guilders

[b] Two-sided *t*-test to assess difference in mean costs between respondents and non-respondents.
SD: Standard deviation

DISCUSSION

The net response rate to the mailed health survey was 70.4%. This is quite satisfactory for a postal questionnaire. A mailed health survey conducted among a general population in the southern part of The Netherlands attained a comparable response rate.[16] However, data from the 29.6% non-respondents are missing. As long as these data are missing in a random fashion the non-response is no

problem. When selective non-response occurs this can affect the conclusions of the health survey. The main conclusion from the present study is that non-response bias will result in a small overestimation of the utilization of out-patient care.

The results of this study showed that response was associated with age, sex, degree of urbanization and type of insurance. Women aged 35-74 years were more likely to participate in the survey than younger women. Men in the younger groups aged 15-54 years were relatively less often among the respondents, while relatively more men aged 65-74 years participated in the survey. Higher response rates occurred in rural areas. Among the disabled and other policy holders the response rates were lower than among employed policy holders. The results with regard to age and degree of urbanization are consistent with other studies[14]. Maybe young and healthy persons, in particular men, are less interested in health, health care and related topics. For them other issues such as their career are possibly more important. This can be an explanation for the relative low response rate among young men. In highly urbanized regions social problems such as unemployment and bad housing conditions accumulate. Persons living in such areas are less willing to participate in a health survey. They probably do not see how they can benefit by completing a questionnaire.

Respondents and non-respondents differ in their utilization of several types of care. Relatively more users than non-users of out-patient specialist care, prescribed drugs, obstetrics, paramedical services, maternity care, technical aids and dental care responded. Since the demographic variables were included in the logistic regression model these results were corrected for differences in these variables. In order to examine the quantity of the utilization, the mean costs of respondents and non-respondents for the different types of care were compared. Respondents had significantly higher mean costs than non-respondents for obstetrics, maternity care, paramedical services and dental care. Although respondents had a higher medical consumption, response was not associated with the utilization of care related to severe conditions. Respondents and non-respondents did not differ in the utilization of in-patient hospital care and transport by ambulances. In a study where the early and late respondents to a postal health survey were compared, the early respondents were described as the 'worried well', healthy individuals who see their doctor regularly, receive disease detection screening and have healthy lifestyles. Early and late respondents did not differ in the rates of death or hospitalization.[17] In another study, persons with ambulatory physician visits were significantly more likely to respond to a survey.[11] In the present study a description such as 'worried well' can be applied to the respondents. The health survey data overestimated the utilization of out-patient care. It is important to

keep this in mind when a postal health survey is used to gather information on the health care use of a population in order to estimate the quantity of health services needed in an area.
Using administrative data on health care expenditures to estimate the utilization of out-patient specialist care results in higher proportions of users of this type of care than using survey data.[18] Therefore the proportion of persons with utilization of out-patient specialist care in table A.3 is larger than the proportion which are usually reported using survey data.[19]
The proportion of users of, for example, dental care among respondents is approximately 8 percentage points higher than among non-respondents. For out-patient specialist care and for paramedical services these figures are respectively 6.5 and 3 percentage points. However, compared to the overall proportions the respondents differ for dental care only 2.5 percentage points and for out-patient specialist care and paramedical services respectively 2 and 1 percentage points. Direct standardization using demographic variables hardly diminishes these differences. The present study shows that when a mailed health survey is used to measure medical consumption in terms of the proportion of the population consulting a specialist and using dental care, prescribed drugs or paramedical services such as physiotherapy, non-response bias will result in a small overestimation of utilization.

REFERENCES

1. Forthofer RN. Investigation of nonresponse bias in NHANES II. Am J Epidemiol 1983;117:507-15.

2. Herzog AR, Rodgers WL. Age and response rates to interview sample surveys. J Gerontol 1988;43:s200-5.

3. Adams MME, Scherr PA, Branch LG et al. A comparison of elderly participants in a community survey with nonparticipants. Public Hlth Rep 1990;105:617-22.

4. Jay GM, Liang J, Liu X, Sugisawa H. Patterns of nonresponse in a national survey of elderly Japanese. J Gerontol 1993;48:s143-52.

5. Colsher PR, Wallace RB. Data quality and age: health and psychobehavioral correlates of item nonresponse and inconsistent responses. J Gerontol 1989;44:p45-52.

6. Sonne-Holm S, Sorensen TIA, Jensen G, Schnohr P. Influence of fatness, intelligence, education and sociodemographic factors on response rate in a health survey. J Epidemiol Commun Hlth 1989;43:369-74.

7. Siemiatycki J, Campbell S. Nonresponse bias and early versus all responders in mail and telephone surveys. Am J Epidemiol 1984;120:291-301.

8. CriQui MH, Barrett-Connor E, Austin M. Differences between respondents and non-respondents in a population-based cardiovascular disease study. Am J Epidemiol 1978;108:367-72.

9. Macera CA, Jackson KL, Davis DR, Kronenfeld JJ, Blair SN. Patterns of non-response to a mail survey. J Clin Epidemiol 1990;43:1427-30.

10. Osler M, Schroll M. Differences between participants and non-participants in a population study on nutrition and health in the elderly. Eur J Clin Nutr 1992;46:289-95.

11. Grotzinger KM, Stuart BC, Ahern F. Assessment and control of nonresponse bias in a survey of medicine use by the elderly. Med Care 1994;32:989-1003.

12. Dillman DA. Mail and telephone surveys: the total design method. New York: Wiley, 1978.

13. Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley, 1989.

14. Den Dulk CJ, Van der Stadt H, Viegen JM. Een nieuwe maatstaf voor stedelijkheid: de omgevingsadressendichtheid [A new measure for degree of urbanization: the address density of the surrounding area]. Maandbericht Bevolking (CBS) 1992;7:14-27.

15. Rowland ML, Forthofer RN. Adjusting for nonresponse bias in a health examination survey. Public Hlth Rep 1993;108(3):380-6.

16. Mackenbach JP, Mheen H, Stronks K. A prospective cohort study investigating the explanation of socio-economic inequalities in health in The Netherlands. Soc Sci Med 1994;38(2):299-308.

17. Paganini-Hill A, Hsu G, Chao A, Ross RK. Comparison of early and late respondents to a postal health survey questionnaire. Epidemiology 1993;4(4):375-9.

18. Lamers LM. Validating survey data via comparison with administrative data. Proceedings of the International Conference on Survey Measurement and Process Quality, Contributed Papers, 1995 April 1-4; Bristol, U.K.. Alexandria: American Statistical Association, 1995:1-5.

19. Swinkels H. Trendcijfers gezondheidsenquête; medische consumptie, 1981-1994 [Trend figures Netherlands Health Interview Survey; medical consumption, 1981-1994]. Maandbericht gezondheidsstatistiek (CBS) 1995;14(5):4-28.

APPENDIX B

Validating survey data on medical consumption via comparison with administrative data [23]

[23] This paper is presented at the International Conference on Survey Measurement and Process Quality, Bristol, U.K., April 1-4, 1995 and published in the proceedings of this conference (Proceedings of the international conference on survey measurement and process quality, American Statistical Association, Alexandria, 1-5.).

SUMMARY

In February 1993 a mail health survey was conducted among about 13,500 adults between 15 and 90 years old and among parents of about 1,500 children aged 5 to 15 years. The final response rate for adults was 70.0% and for children 75.4%. The questionnaire contained questions about health and medical consumption. All participants were enroled with one health insurance company and were included in a panel data set that could be matched with the health survey data. For each participant the panel data set comprises administrative information on hospitalizations, annual health care expenditures and demographic variables. The answers to the questions about contact with a specialist, a physiotherapist, a speech therapist and hospital admissions during the last year could be compared with administrative data from 1992. Although there is a small time lag the agreement between survey and administrative data was high for contact with a physiotherapist and a specialist, for both adults and children. For contact with a speech therapist and hospital admissions the agreement between the data sources was low. The low agreement for contact with a speech therapist can be explained by incomplete administrative data; for hospital admissions by recall bias resulting in an overestimation of admissions.

INTRODUCTION

Health surveys often contain questions like: 'During the past 12 months, how many times did you visit a (medical) specialist?' or 'During the last 12 months, how many hospitalizations did you have?'. Recall bias can distort the answers to such questions. To assess the validity of self-reports, survey data are compared with other data sources, containing comparable information. Several studies mention comparisons between surveys, either mailed, telephone or face to face, with medical records (Harlow and Linet[1] give an overview). Most studies comparing questionnaires with medical records examine (chronic) medical conditions. The agreement between the data sources varies per condition[2-7].

A few studies made comparisons with administrative or claims data. Carsjo et al.[8] reported that among the very old the agreement between utilization versus no utilization during the past three months was very high for hospitalization and substantial for visits to physicians. Glandon et al.[9] have found that elderly persons underreport physician utilization. The total discrepancy between archival and self-reported measures was nearly two visits per person. According to Berk et al.[10] elderly persons are not able to accurately recall their expenditures for prescription drugs.
This study compares self-reported data from respondents to a mailed health survey with administrative data on medical consumption.

DATA

In February 1993 a first mailing, containing the questionnaire with the cover letter, was sent. After one week everyone received a postcard reminder, which served both as a thank you for those who had already responded and as a friendly reminder for those who had not. Three weeks after this postcard reminder a letter with a new questionnaire was sent to the nonrespondents only. Two weeks after this, the nonrespondents received a final mailing, containing a letter.

The main purpose of the survey was to gather information on health status and (additional) medical consumption to be used for a study on risk-adjusted capitation payments to health insurers, an important issue in the Dutch health care reforms. About 15,000 people received the health survey. Among them were 13,472 adults between 15 and 90 years old and parents of 1,509 children aged 5 to

14 years. The parents were asked to complete the questionnaire for their child. All participants were enrollees from one social health insurer, a so-called sickness fund, in the western part of the Netherlands and they were included in a panel data set that could be matched with the health survey data. For each participant the panel data set comprises for five years, from 1988 to 1992, administrative information on hospitalizations (when applicable), annual health care expenditures and demographic variables such as age, sex, zip-code and type of insurance.

Survey data were compared with these administrative data for date of birth and sex to make sure that the eligible person and not someone else completed the questionnaire. For the adults 413 questionnaires were completed by another person; 64 parents completed the questionnaire for another child than was asked for. Questionnaires completed by other than the eligible person are considered as nonresponse, resulting in a net response rate for adults of 70.0% and for children of 75.4% The net response rate for the total sample is 70.4%.

The questionnaire contained questions about health and medical consumption. The health indicator questions refer to perceived health status, chronic conditions, functional disabilities and vague psycho-somatic complaints. The respondents were also asked some demographic information, like date of birth and sex. The questionnaire further contained questions about the respondent's country of birth and that of his parents, education, marital status and the number of persons in the household.

The questions about medical consumption refer to the consultation of a general practitioner during the last two months and to the consultation of a specialist, physiotherapist, speech therapist and dentist during the last 12 months. The respondents were also asked how many times they were hospitalized during the last 12 months and if they used prescribed drugs during the last 14 days.

METHOD

The answers to the questions about contact with a specialist, a physiotherapist and a speech therapist (questionnaire for children only) and about hospital admissions during the last year could be compared with administrative data from 1992. When someone answered that he went to a physiotherapist last year, we

should find costs for physiotherapy in the administrative data. The comparison for using speech therapy is only made for children and not for adults, because of the (very) low prevalence of speech therapy for adults (0.5%). No comparison for consultation of a general practitioner could be made, because GPs in the Netherlands receive a uniform annual fee for each of the patients on their list that is enroled with a sickness fund regardless of medical consumption. No comparison for visiting a dentist is made, because Dutch children younger than 18 years are obliged by the sickness fund to visit a dentist twice a year. When they fail to do so, the parents have to pay for their dental care out of pocket.

Predictive values were calculated for both a positive and a negative answer to the survey questions. The predictive value of a positive answer to the question about consultation of a specialist is the probability that for a respondent with the answer 'yes, I consulted a specialist' there are costs for care provided by a specialist in the administrative data. The predictive value of a negative answer to the question about consultation of a specialist is the probability that for a respondent who answered no consultation of a specialist, the costs for care provided by a specialist are zero.
Finally, kappa values are calculated as a measure of agreement between the two data sources.

RESULTS

The percentage of adult respondents consulting a specialist is higher for the administrative data (44.9%) than for the survey data (38.9%). The prevalence of people hospitalized in one year is higher for the survey data than for the administrative data, 10.2% versus 6.9%. The percentages of respondents using physiotherapy is a little higher for the survey data than for the administrative data (table B.1).

For the respondents with costs for specialist care, who said not to have consulted a specialist, the administrative records were checked for the medical specialisms concerned. In 37% of these cases it concerned radiology, radiography or radiotherapy; in 25% of the cases consultation of an eye (ophthalmic) specialist; in 20% surgery; all other specialisms appeared less than 10%. Apparently respondents did not consider a visit to a hospital for radiotherapy or radiography as a consulta-

tion of a specialist. Dutch eye specialist often work in private clinics. However, for a lot of people a specialist is a (specialized) doctor working in the hospital. The lowest predictive value for the negative answers is found for the consultation of a specialist. For the positive answers the predictive value for hospital admission is low.

Table B.1 Comparison of survey data with administrative data for adults

	relevant costs in 1992		
survey question	yes	no	predictive value
contact with specialist			
positive answer	32.5 %	6.4 %	84 %
negative answer	12.4 %	48.7 %	80 %
hospital admission			
positive answer	6.1 %	4.1 %	59 %
negative answer	0.8 %	89.0 %	99 %
contact with physiotherapist			
positive answer	14.7 %	5.0 %	75 %
negative answer	3.6 %	76.7 %	96 %

The results for children are comparable to those for adults. Table B.2 shows the results for the children. The percentage of children consulting a specialist is a little higher for the administrative data (33.6%) than for the survey data (31.9%). The percentage of children using physiotherapy is almost the same for the two sources of data. The survey data overestimated the percentage of children hospitalized and the percentage of children consulting a speech therapist, resulting in low predictive values for the positive answers.

With the question concerning hospital admissions was implicitly meant a stay in the hospital during at least one night (inpatient care). However, a night in the hospital can be a traumatic experience, especially for children. Therefore, children are treated as much as possible in day case treatment or short stay, hospital admissions for four to eight hours, (e.g. for tonsillectomy). Day case treatment is a form of outpatient care, care provided in an ambulatory setting. If the costs of day case treatment are included in the costs for hospitalization the results of the comparison for hospitalizations change. Table B.3 shows that for children the prevalence of hospitalizations from administrative data becomes almost the same

as for the survey data. Parents apparently gave affirmative answers to the question about hospital admission when their child had day case treatment. For the adults the prevalence of hospitalizations from administrative data increased from 6.9% to 8.6%, which is still less than the 10.2% from the survey data.

Table B.2 Comparison of survey data with administrative data for children

survey question	relevant costs in 1992 yes	relevant costs in 1992 no	predictive value
Contact with specialist			
positive answer	24.7 %	7.2 %	77 %
negative answer	8.9 %	59.2 %	87 %
Hospital admission			
positive answer	2.1 %	5.4 %	29 %
negative answer	0.3 %	92.2 %	100 %
Contact with physiotherapist			
positive answer	2.8 %	1.0 %	73 %
negative answer	0.7 %	95.5 %	99 %
Contact with speech therapist			
positive answer	3.1 %	6.7 %	32 %
negative answer	0.3 %	89.9 %	100 %

Table B.3 Comparison of survey question concerning hospital admission with administrative data, including day case treatment for adults and children

survey question	relevant costs in 1992, including day case treatment yes	relevant costs in 1992, including day case treatment no	predictive value
Adults: hospital admission			
positive answer	7.3 %	2.9 %	71 %
negative answer	1.3 %	88.5 %	99 %
Children: hospital admission			
positive answer	6.1 %	1.4 %	81 %
negative answer	1.6 %	90.9 %	98 %

The survey started in february 1993. The respondents where asked questions about the last 12 months, resulting in a small time lag between the survey and the administrative data. When we took into account only the cost from March 1992 until the end of the year for the comparison with survey data, the results were comparable to the results presented here.

Table B.4 shows the Kappa statistic, which is calculated as a measure of agreement between the data sources. According to Landis and Koch[11] for kappa's between .40 and .60 the agreement can be classified as fair; for kappa's between 0.60 and 0.80 as substantial. As could be expected the kappa for using speech therapy and for hospital admission for children is relative low. Including day case treatment in costs for hospitalizations improved the kappa, especially for children. The agreement between the data sources for consulting a specialist and using physiotherapy can be qualified as substantial.

Table B.4 Kappa for children and adults

	children	adults
Contact with specialist	0.63	0.62
Hospital admission	0.42	0.68
(including day case treatment)	(0.78)	(0.78)
Contact with physiotherapist	0.75	0.72
Contact with speech therapist	0.45	

DISCUSSION

For consulting a specialist the agreement between the survey and the administrative data is good. However, the prevalence estimates for the administrative data are higher than for the survey data. This discrepancy is in many cases caused by respondents who according to the administrative data visit the hospital for radiology, radiography or radiotherapy or consulted an eye specialist and gave a negative answer to the question about consultation of a specialist.

The agreement between the data sources for using physiotherapy is good. The agreement between the data sources for speech therapy is low. The overreporting of speech therapy by the survey data can be explained by the fact that many speech therapists in the Netherlands are employees of schools or hospitals. The sickness

fund only reimburses speech therapists working in private clinics. So, the administrative data contained only the costs of these speech therapists.

For hospital admissions the survey data gave higher prevalence estimates than the administrative data. For children this can be explained by the fact that parents gave affirmative answers when the child went to the hospital for day case treatment. Confusion of day case treatment in the hospital with hospitalizations, as inpatient care, is for adults only partly an explanation for the differences in prevalences between the two data sources. Another explanation is forward telescoping. In forward telescoping, the respondent includes events from a previous time period in the period asked about[12]. This form of telescoping is likely to occur for highly salient topics such as hospitalizations.

REFERENCES

1. Harlow S.D., M.S. Linet. Agreement between questionnaire data and medical records. The evidence for accuracy of recall. American Journal of Epidemiology 1989; 129: 233-48.

2. Bush T.L., S.R. Miller, A.L. Golden, W.E. Hale. Self-Report and Medical Record Report Agreement of Selected Medical Conditions in the Elderly. American Journal of Public Health 1989; 79: 1554-6.

3. Heliovaara M., A. Aromaa, T. Klaukka e.a.. Reliability and validity of interview data on chronic diseases. Journal of Clinical Epidemiology 1993; 46: 181-91.

4. Linet M.S., S.D. Harlow, J.K. McLaughlin, L.D. McCaffrey. A comparison of interview data and medical records for previous medical conditions and surgery. Journal of Clinical Epidemiology 1989; 42: 1207-13.

5. Midthjell K., J. Holmen, A. Bjorndal, P.G. Lund-Larsen. Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trondelag diabetes study. Journal of Epidemiology and Community Health 1992; 46: 537-42.

6. Olsson L., K. Svärdsudd, G. Nilsson e.a.. Validity of a postal questionnaire with regard to the prevalence of myocardial infarction in a general population sample. European Heart Journal 1989; 10: 1011-16.

7. Schrijvers C.T.M., K. Stronks, D.H. van de Mheen e.a.. Validation of Cancer Prevalence Data from a Postal Survey by Comparison with Cancer Registry Records. American Journal of Epidemiology 1994; 139: 408-14.

8. Carsjo K., M Thorslund, B. Wärneryd. The validity of survey data on utilization of health and social services among the very old. Journal of Gerontology: Social Sciences 1994; 49: S156-64.

9. Glandon G.L., M.A. Counte, D. Tancredi. An Analysis of Physician Utilization by Elderly Persons: Systematic Differences Between Self-Report and Archival Information. Journal of Gerontology 1992; 47: s245-s252.

10. Berk M.L., C.L. Schur, P. Mohr. Data Watch: Using Survey Data to Estimate Prescription Drug Costs. Health Affairs 1990; 9: 146-56.

11. Landis J.R., G.G. Koch. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-74.

12. Sudman S., N.M. Bradburn. Asking Questions. San Francisco, Jossey-Bass Inc., 1987.

APPENDIX C: Diagnostic Cost Group (DCG) classifications by ICD-9-CM Diagnostic SubGroup (DSG)

DSG [a]	ICD-9-CM code		DCG Ash et al.	DCG present study
1	618	Genital prolapse	1	1
2	630-676	Pregnancy-related problems		
	760-779		1	1
3	610-611	Disorders of breast	1	1
4	540-542	Appendicitis, excluding other disease	1	1
5	380-389	Diseases of ear and mastoid processes	1	1
6	603-609	Diseases of male genitalia except prostate	1	1
7	574-576	Cholelithiases and other disorders of gallbladder and biliary tract	2	1
8	550-553	Hernia of abdominal cavity	2	1
9	470-478	Other diseases of respiratory tract	2	1
10	210-229	Benign neoplasms	2	1
11	600-602	Prostate disorders	2	2
12	366	Cataract	2	2
13	350-359	Disorders of peripheral nervous system	3	3
14	598	Urethral stricture	3	2
15	617	Other diseases of female genital tract,		
	619-629	excluding prolapse, abnormal bleeding	3	1
16	360-364	Disorders of the eye and adnexa, except for		
	367-379	cataract and glaucoma	3	1
17	365	Glaucoma	3	1
18	591-594	Hydroephrosis, calculus of kidney and ureter, other disorders of kidney and ureter, calculus of lower urinary tract	3	3
19	800-839	Injuries involving fractures and dislocations	3	1
20	726-729	Rheumatism, excluding the back and polymyalgia rheumatica	3	1
21	721-724	Dorsopathies, except for inflammatory spondylopathies	3	1
22	590	Kidney infections	3	1

DSG [a]	ICD-9-CM code		DCG Ash et al.	DCG present study
23	910-919	Superficial injury		
	920-924	Contusions		
	930-939	Effects of object entering though orifice	3	1
24	780	General symptoms	3	3
25	850-909	Intracranial injury (excluding skull fracture), internal injury, open wounds, injury to blood vessels, late effects of injuries, poisonings, toxic effects and external causes		
	925-929	Crushing injury		
	940-949	Burns		
	950-959	Injury to nerves, spinal cord, unspecified injuries	3	1
26	520-529	Diseases of oral cavity, salivary glands, and jaws	3	1
27	460-465	Acute respiratory infections, except bronchitis	3	1
28	555-556	Noninfective enteritis and colitis		
	558		4	3
29	740-759	Congenital anomalies	4	1
30	840-848	Sprains and strains of joints and muscles	4	1
31	170-171	Malignant neoplasm of bone, skin, cartilage,		
	173-175	soft tissue, male and female breasts		
	233			
	172	Skin melanoma and carcinoma	4	1
32	560-562	Intestinal obstruction (nonherniated), diverticula of intestine		
	567-569	Peritonitis and other disorders of intestine or peritoneum	4	3
33	595-597	Disorders of urethra and urinary tract		
	599			
	788,791	Urinary symptoms and nonspecific findings on urine examination	4	2
34	530-534	Gastric, duodenal, peptic, and gastrojejunal ulcer, diseases of the esophagus		
	578	Gastrointestinal haemorrhage	4	3
35	451-459	Diseases of veins and lymphatics and diseases of circulatory system		
	565-566	Anal fissures, fistulae and anal or rectal abscess	4	1
36	001-139	Infectious diseases, except those in DSG 71	4	1

DSG [a]	ICD-9-CM code		DCG Ash et al.	DCG present study
37	535-537	Gastritis and duodenitis and other disorders of stomach and duodenum		
	564	Functional digestive disorders		
	787	Digestive symptoms	4	3
38	251-254	Diseases of other endocrine glands		
	256-257			
	259			
	240-246	Disorders of the thyroid gland		
	783	Nutritional and metabolic symptoms		
	790	Nonspecific findings on blood examinations	4	1
39	430-438	Cerebrovascular disease	4	3
40	786	Respiratory symptoms	4	2
41	480-487	Pneumonia and influenza	4	2
42	680-709	Diseases of the skin and subcutaneous tissue		
	782	Symptoms involving skin	4	2
43	789	Symptoms involving abdomen and pelvis	4	2
44	710	Diseases of connective tissue		
	714	Rheumatoid arthritis and inflammatory polyarthropathies		
	715	Osteoarthrosis and like disorders		
	390-398	Rheumatic fever and rheumatic heart diseases		
	446	Polyarthritis and like conditions		
	720	Inflammatory spondylopathies		
	725	Polymyalgia rheumatica		
	781	Symptoms of nervous and musculoskeletal systems	5	3
45	711-713	Various arthropathies		
	716			
	717-719	Disorder and derangement of joints		
	730-739	Osteopathies, chondropathies, and acquired musculoskeletal deformities	5	2
46	614-616	Inflammatory disease of female pelvic organs	4	1
47	410-414	Ischemic heart disease		
	415-417	Diseases of pulmonary circulation		
	785	Cardiovascular symptoms	5	2
48	441	Aortic aneurysm		
	442	Other aneurysms		
	444	Arterial embolism and thrombosis	5	2

DSG [a]	ICD-9-CM code		DCG Ash et al.	DCG present study
49	300-302	Neurotic, personality, and sexual disorders		
	303-305	Alcohol and drug dependence and abuse		
	306-319	Other personality disorders, nonpsychotic mental disorders, and mental retardation	5	1
50	792-796	Nonspecific abnormal findings		
	799	Other ill-defined and unknown causes of morbidity and mortality	0	0
51	V1-V82	Supplementary classification of factors influencing health status	5	2
52	466	Acute bronchitis and bronchiolitis	5	2
53	290-294	Organic psychotic conditions		
	295-299	Other psychoses		
	797	Senility without psychosis	5	1
54	577,579	Diseases of pancreas, intestinal malabsorption	5	3
55	401-405	Hypertensive disease	5	3
56	440	Atherosclerosis		
	443	Other vascular disease		
	447-448	Disorders of arteries, arterioles, and capillaries		
	557	Vascular insufficiency of intestine	5	3
57	260-269	Nutritional deficiencies		
	270-275	Metabolic disorders		
	277			
	278-279	Obesity and immune disorders		
	280-289	Disorders of blood and blood-forming organs		
	255,258	Adrenal and polyglandular disorders	5	2
58	232,234	Carcinoma in situ in skin and other nonspecified sites		
	235-239	Neoplasms of uncertain behavior or unspecified nature		
	140	Malignant neoplasm of the lip	5	2
59	276	Disorders of fluid, electrolyte, and acid base balance	5	3
60	420-427	Other forms of heart disease, except heart		
	429	failure	5	3
61	980-989	Toxic effects of nonmedical substances		
	990-995	Unspecified effects of external causes	5	1

DSG [a]	ICD-9-CM code		DCG Ash et al.	DCG present study
62	340-349	Other disorders of the central nervous system	5	3
63	960-979	Poisoning by drugs, medicines, and biological substances	5	1
64	330-337	Hereditary and degenerative diseases of the central nervous system	5	3
65	996-999	Complications of medical care not elsewhere classified	5	3
66	784	Symptoms involving head and neck	5	2
67	250	Diabetes mellitus	6	3
68	150-159	Malignant neoplasm of digestive organs and peritoneum		
	230	Carcinoma of digestive organs	6	1
69	179-189	Malignant neoplasm of genitourinary organs	6	1
70	570-573	Liver disorders and diseases	7	3
71	013,038	Various infections and parasitic diseases		
	045-049			
	070,112			
	093-095			
	114-116			
	135			
	320-326		7	2
72	285	Other and unspecified anemias	7	3
73	490-496	Chronic obstructive pulmonary disease and like conditions		
	500-508	Pneumoconioses and other lung diseases due to external agents	7	4
74	510-519	Other diseases of the respiratory system	7	3
75	428	Heart failure	8	4
76	141-149	Malignant neoplasm of oral cavity and pharynx		
	190-199	Malignant neoplasm of unspecified sites		
	200-208	Malignant neoplasm of lymphatic and haematopoietic tissue	8	3
77	160-165	Malignant neoplasm of respiratory and intrathoracic organs	8	1
78	580-589	Nephritis, nephrotic syndrome, and nephrosis	9	5

[a] The diagnostic subgroups classification was developed by Ash et al. (1989).

APPENDIX D: Mean follow-up costs per Diagnostic SubGroup (DSG)

DSG	Major Diagnostic Category [a]	number of 1988 admissions	follow-up costs 1989	age/sex corrected follow-up costs 1989	follow-up costs 1990 + 1991 [b]
1	Kidney and urinary tract, reproductive system	135	1816	718	2331
2	Pregnancy	1650	1093	451	1385
3	Kidney and urinary tract, reproductive system	41	847	211	1506
4	Digestive system	237	755	252	785
5	Nervous system and sense-organs	243	1125	556	1075
6	Kidney and urinary tract, reproductive system	183	1019	369	1675
7	Digestive system	293	1745	719	1696
8	Digestive system	531	1457	481	1680
9	Respiratory system	890	663	198	832
10	Neoplasm	140	1191	527	1053
11	Kidney and urinary tract, reproductive system	243	2477	1103	3168
12	Nervous system and sense-organs	500	2055	459	2994
13	Nervous system and sense-organs	96	4335	3494	4620
14	Kidney and urinary tract, reproductive system	82	1905	992	2522
15	Kidney and urinary tract, reproductive system	577	1483	734	1560
16	Nervous system and sense-organs	316	2017	1309	1472
17	Nervous system and sense-organs	42	1908	514	1929
18	Kidney and urinary tract, reproductive system	177	3875	3061	4291
19	Injuries, poisonings, burns	677	2572	1643	2208
20	Musculoskeletal system and connective tissue	382	1799	1068	1752
21	Musculoskeletal system and connective tissue	813	1886	1193	2216
22	Kidney and urinary tract, reproductive system	15	1520	972	1884
23	Injuries, poisonings, burns	56	2343	1570	2170

Table continued

DSG	Major Diagnostic Category [a]	number of 1988 admissions	follow-up costs 1989	age/sex corrected follow-up costs 1989	follow-up costs 1990 + 1991 [b]
24	Symptoms	344	3790	2845	3695
25	Injuries, poisonings, burns	374	2027	1387	1123
26	Digestive system	159	1568	942	1511
27	Respiratory system	32	2152	1648	1439
28	Digestive system	115	3774	3003	4977
29	Congenital anomalies	176	1646	1227	1786
30	Injuries, poisonings, burns	55	1020	465	918
31	Neoplasm	171	2574	1642	1927
32	Digestive system	210	3585	2562	3432
33	Kidney and urinary tract, reproductive system	243	2336	1452	2874
34	Digestive system	220	3613	2483	3725
35	Circulatory system	505	1504	670	1779
36	Infectious and parasitic diseases	96	2717	1972	2288
37	Digestive system	133	3906	3201	4253
38	Endocrine, nutritional and metabolic diseases	187	2889	2059	2075
39	Circulatory system	229	4782	3540	3404
40	Symptoms	243	3791	2930	3100
41	Respiratory system	155	2727	1965	2740
42	Skin, subcutaneous tissue	343	2859	2041	2837
43	Symptoms	487	2491	1685	2820
44	Musculoskeletal system and connective tissue	280	4765	3650	3627
45	Musculoskeletal system and connective tissue	942	2375	1565	2625
46	Kidney and urinary tract, reproductive system	49	2220	1662	1892
47	Circulatory system	1175	3750	2630	3042
48	Circulatory system	116	3364	2479	2854
49	Mental diseases, alcohol and drug abuse	332	3894	3292	1960
50	Symptoms	5	8576	7478	2275
51	Factors influencing health status	1273	2821	2133	2773

Table continued

DSG	Major Diagnostic Category [a]	number of 1988 admissions	follow-up costs 1989	age/sex corrected follow-up costs 1989	follow-up costs 1990 + 1991 [b]
52	Respiratory system	2	296	75	1011
53	Mental diseases, alcohol and drug abuse	255	2524	1766	2317
54	Digestive system	45	2992	2212	4228
55	Circulatory system	48	2073	1071	4664
56	Circulatory system	130	4028	3129	3865
57	Endocrine, nutritional and metabolic diseases	103	1561	843	3020
58	Neoplasm	458	1978	1078	2629
59	Endocrine, nutritional and metabolic diseases	32	3588	2844	3522
60	Circulatory system	266	3232	2256	3553
61	Injuries, poisonings, burns	71	1591	1049	1692
62	Nervous system and sense-organs	224	5267	4571	3457
63	Injuries, poisonings, burns	55	3288	2745	1896
64	Nervous system and sense-organs	73	4401	3430	5031
65	Injuries, poisonings, burns	112	3824	3197	4741
66	Symptoms	81	2410	1386	3099
67	Endocrine, nutritional and metabolic diseases	233	5214	4092	4948
68	Neoplasm	60	3471	2375	1848
69	Neoplasm	109	3548	2655	2090
70	Digestive system	45	3173	2336	3805
71	Infectious and parasitic diseases	71	1735	1170	2831
72	Endocrine, nutritional and metabolic diseases	70	7073	5848	3841
73	Respiratory system	277	5611	4633	6606
74	Respiratory system	169	3257	2418	3479
75	Circulatory system	161	5696	4450	6400
76	Neoplasm	94	3861	2969	3781
77	Neoplasm	44	3012	2040	2282
78	Kidney and urinary tract, reproductive system	20	12575	11879	30701

[a] Appendix C gives an overview of ICD-9-CM codes per DSG.

[b] Before adding costs in 1990 plus 1991, costs in 1990 and 1991 were weighted in order to make the mean costs in these years equal to the mean costs in 1989.

Curriculum vitae

After her graduation from the Sint Oelbert Gymnasium in Oosterhout, Leida Maria Lamers (1962) studied psychology at the University of Amsterdam. In 1987 she gained her masters degree (cum laude). During her study she assisted in a research methods training program (1985-1988). From 1988 to 1989 she was a researcher at the department of General Practice at the University of Leiden. At the department of Epidemiology of the municipal Public Health Service of Rotterdam she worked as a statistician from 1989 to 1992. In 1992 she joined the department of Health Policy and Management at the Erasmus University Rotterdam. Her research focuses on the use of diagnostic information to adjust capitation payments to competing Dutch sickness funds.

Acknowledgements

In 1992 Wynand van de Ven and René van Vliet initiated the research on risk-adjusted capitation payments that resulted in this thesis. Although they initially were looking for an econometrist, I am glad they gave me, a psychologist, the opportunity to perform the study. Thanks to both of them for the stimulating discussions on risk-adjustment and related topics and their useful criticisms on the drafts of reports, papers and this thesis. I am indebted to René for sharing his knowledge on statistics and SAS with me.

My colleagues at the department of Health Policy and Management provided a stimulating working environment. Thanks. Two persons I would like to mention. Frank Bakker with whom I shared a room for two years and never complained about the temperature in our office. After the department moved to 'Woudestein' Frank and Ad Vermaas became my neighbours. The statement 'better a good neighbour than a far friend' applied to both of you. Thanks for all your informal help.
I gratefully acknowledge the assistance of Barrie van Gameren and Fiona Buruma with the computations for this study.

I am grateful to VNZ (the Association of Dutch Sickness Funds, currently merged into Zorgverzekeraars Nederland) for financial support and to the health insurance organization 'Zorg en Zekerheid' for providing the data set. I would like to thank Rick Ouwerkerk, Toine Slaats, Frans Trapman and Jeroen van der Meulen for successively answering all my questions about the numerous data files they sent to me over the years. The members of the 'begeleidingscommissie Budgettering en Zorgverzekering' Ron Hendriks, Wim Nuyens †, Jan Collaris, Han van der Knoop and Wim van Tol are gratefully acknowledged for their comments on drafts of reports, which preceded this thesis. Thanks also to all the anonymous members of Zorg en Zekerheid who responded to the health survey.

Special thanks to Tanja Algra for visualizing the process of risk-adjustment. She designed the beautiful cover.

Finally a word of thanks to my family, friends and, of course, Dolf for their support. Life is so much more fun with you around.