Behrend, Corinne; Buchner, Florian; Happich, Michael; Holle, Rolf; Reitmeir, Peter; Wasem, Jürgen

Working Paper

Risk-Adjusted Capitation Payments: How Well Do Principal Inpatient Diagnosis-Based Models Work in the German Situation? Results From a Large Data Set

Diskussionsbeiträge aus dem Fachbereich Wirtschaftswissenschaften der Universität Duisburg-Essen, Campus Essen, No. 134

Provided in cooperation with:
Universität Duisburg-Essen (UDE)

Diskussionsbeiträge aus dem
Fachbereich Wirtschaftswissenschaften
Universität Duisburg-Essen
Campus Essen

Nr. 134

Mai 2004

Risk-Adjusted Capitation Payments: How Well Do Principal Inpatient Diagnosis-Based Models Work in the German Situation? Results From a Large Data Set.

Behrend Corinne, Buchner Florian, Happich Michael, Holle Rolf,
Reitmeir Peter, Wasem Jürgen

Inhaltsverzeichnis

Abstract

The Risk Adjustment Reform Act of 2001 mandates that a health-status-based risk adjustment mechanism has to be implemented in Germany's Statutory Health Insurance system by January 1, 2007. German parliament decided this as with the existing demographic risk adjustment model, that means there is cream skimming and sickness funds hesitate to engage in managing care for the chronical ill.

Four approaches were used to test the feasibility of incorporating use of diagnosis *as a proxy measure for health status in a German risk adjustment formula. The first two models used standard demographic and socio-demographic variables. The other two models are separately incorporating a simple binary indicator for hospitilization and Hierarchical Coexisting Conditions (HCCs: DxCG® Risk Adjustment Software Release 6.1) using inpatient diagnosis.*

Age and gender grouping accounted for 3.2% of the variation in total expenditures for concurrent as well as prospective models. The current German risk adjusters age, sex, and invalidity status account for 5.1% and 4.5% of the variance in the concurrent and prospective models respectively. There are substantial increases in explanatory power, however, when HCCs are added. Age, gender, invalidity status and HCC covariates explain about 37% of the variations of the total expenditures in a concurrent model and roughly 12% of the variations of total expenditures in a prospective model. For high-risk (cost) groups, substantial under-prediction remains; conversely, for the low-risk group, represented by enrolees who did not show any health care expense in the base year, all of the models over-predict expenditure.

Key words: Risk Adjustment, HCCs, Germany

1. Introduction

In 1993, Germany adopted new Statutory Health Insurance (SHI) legislation to grant all enrolees free choice of health insurer and promote competition among sickness funds. The motive behind this move from an originally captive employment-based social insurance system to a system of a competitive insurance market has been to increase cost responsibility of sickness funds and to secure improvements in terms of efficiency, quality, innovation, and responsiveness to consumer preferences.

About 90% of the German population are offered nearly universal access to health care under largely compulsory and non-profit insurance schemes – the sickness funds –, which, together, make up the SHI. Individuals who are not insured through SHI, mostly civil servants and the self-employed (about 10 per cent or 8.5 million in 2003) carry commercial insurance offered by private health insurance companies. Health care under SHI is based on the notion of solidarity and financed through earnings-related contributions by individuals, with matching employer payments; the insurance cover automatically includes all non-earning dependents (without own income).

The sickness fund market is highly regulated. Open enrolment (under which a sickness fund must, in principle, accept all applicants) and community-rating are required. Benefits under SHI are largely standardised and portable. The package is comprehensive and encompasses preventive, diagnostic, therapeutic and rehabilitative benefits in kind as well as in cash (primarily sickness benefits). This increases incentives for sickness funds to encourage enrolment of better health risks and discourage enrolment of worse ones, thereby competing on risk selection and not on price, service, quality and efficiency. Risk selection as well as adverse selection and other types of self-selection contributes to risk segmentation, in which sickness funds experience different levels of risk in the populations they cover. To help the sickness fund market function properly and create a "level playing field", the 1993 legislative reforms also introduced a risk adjustment scheme, which on the one hand adjusts for differences in the income of the insured (as the base for income related contributions) and on the other hand adjusts for expenditure risks of the enrolee: Sickness funds pay an income related solidarity contribution (for the terminology see [1]) into the risk adjustment mechanism, and in return they receive a risk adjusted premium subsidy from that pool.

Different approaches for grouping enrolees and predicting those groups' health care expenses give rise to different risk adjustment systems. In Germany, the premium subsidy which the

risk adjustment pool pays to the sickness funds is primarily adjusted for age, sex and an invalidity status indicator (i.e., the drawing of disability benefits).[1] There is evidence that these socio-demographic factors are much too crude to reflect actual health care expenditures accurately. Only about 5% of the variation in SHI allowances in kind has been found to be explained by age, sex and disability status as risk factors [3]. Concerns about the limitations of the current risk adjustment system resulted in the Risk Adjustment Reform Act of 2001, which mandates that by 2007 the existing system of risk adjustment be replaced with one that takes enrolees' health status into account. International research indeed shows risk adjustment models that include utilisation of health services as a proxy measure for health status – diagnoses, procedures, and prescriptions from administrative claims data for instance – perform much better than systems based on demographics or socio-demographics alone[2].

To date, there has only been limited research on risk assessment and adjustment in Germany. The lack of data is indeed a problem. Current legislative provisions covering privacy as well as other aspects, especially the tradition of self-governance for sickness funds and providers alike and its brassbound apologia , place constraints on the linking of data sets to trace patient encounters within the health system. Several countries, however, notably the US and the Netherlands, have had experience in developing, implementing and refining methods of risk assessment based on health care utilisation data to explain and predict cost variation and hence to set risk-adjusted capitation payments. Some of these health-based risk adjusters, which differ substantially in the "heritage", logic and most common applications, have now been adopted by a number of US payers and the Dutch government. The practical question then becomes whether it is possible to use these same risk adjustment methods in Germany for the purpose of categorising health risks sufficiently well to mitigate the financial rewards and penalties of risk selection and adverse selection.

In this paper we report on the first application and adaptation of a US system of risk assessment to a German sickness fund population. We sought to evaluate the performance of a major diagnosis-based case-mix measure, the Diagnostic Cost Group/Hierarchical Condition Category (DCG/HCC), in explaining variation in health care resource use in a SHI population of North-Eastern Germany. Specifically, we assess the explanatory power of this risk adjustment model in a concurrent (i.e., same-year) and prospective (next-year) framework.

---

[1] The adjustment mechanism also accounts for the type of entitlement to sickness benefits, which are income-related allowances in cash and shall not be considered in this study. For a detailed analysis of the existing risk adjustment mechanism in Germany see [2].

[2] See for an overview [1]

1.1 DCG/HCC risk assessment method

The Diagnostic Cost Group/Hierarchical Condition Category (DCG/HCC) system is a family of diagnosis-based risk profiling and assessment methods developed by Arlene Ash, Randall Ellis and colleagues at Boston University and Lisa Iezzoni of Harvard Medical School. The single-condition DCG models originally sought to predict year-2 expenditure for Medicare beneficiaries 65 years of age and older based on their single worst principal inpatient diagnosis in year 1 [4]. The classification methodology has since been refined, adapted to other populations, and extended to predict year-1 (concurrent modelling) as well as year-2 (prospective modelling) expenditure on the basis of both ambulatory and inpatient diagnoses of year 1. Also, with the more recent multi-condition HCC models, predictions are derived from the full set of medical conditions present from individuals' encounters with the health care system, i.e., the cumulative expenditure effect of multiple conditions is captured. To limit the sensitivity of these models to coding idiosyncrasies and code proliferation, multiple interrelated ICD-9-CM diagnosis codes for common conditions are grouped and arrayed in a hierarchy based on expense; within each hierarchy, individuals are only assigned to the highest group, which stands for their worst, i.e., their most expensive diagnosis .

We gauged the models' relative overall performance by examining the deviation of the intercept and slope estimates. To characterize the health status of individuals, the DCG/hierarchical condition category (HCC) concurrent and prospective model, as implemented in $D_xCG^{®}$ Release 6.1 was used [5].

1.2 Literature in Section "Introduction":

1. van de Ven, W.P.M.M. and R. Ellis, *Risk Adjustment in competitive health plan markets*, in *Handbook of Health Economics*, A.J. Culyer and J.P. Newhouse, Editors. 2000, Elsevier North Holland: Amsterdam. p. 755-845.
2. Buchner, F. and J. Wasem, *Needs for further Improvement: Risk Adjustment in the German health insurance system.* Health Policy, 2003. **65**(1): p. 21-35.
3. Behrend C, et al., *Zur Erklärungskraft des heutigen soziodemographischen Risikostrukturausgleichsmodells - Ergebnisse empirischer Analysen an Prozessdaten einer ostdeutschen Regionalkasse.* Journal of Public Health/ Zeitschrift für Gesundheitswissenschaften, 2004. **12**(20-31).
4. Ash, A., F. Porell, and P. Randall, *Adjusting Medicare capitation payments using prior hospitalization data.* Health Care Financing Review, 1989. **10**(Summer 1989): p. 177-188.
5. DxCG, *Risk Adjustment Software: User's Guide Release 6.1*. 2002, Boston, MA: DxCG.

2. Data and Methods

2.1 Data

This study used administrative data files of a population of over 755,000 individuals of all ages who were insured with a regional sickness fund operating in North-Eastern Germany at any time during a period of two consecutive years, 1997 and 1998. An anonymised unique person identifier allowed for the linkage of data files to create one complete data set containing 1997 and/or 1998 socio-demographic, diagnostic and expenditure information for each insuree.

Available socio-demographic information comprised the date of birth, gender, an indicator for the receipt of an invalidity pension, the length of entitlement to long-term care (LTC) insurance benefits, and the time span of enrolment per calendar year including an indicator for death when applicable. For the concurrent analyses of this article, all persons with any cover in 1997 were retained, i.e., a population of 788,130 individuals. Out of the 755,926 individuals with any cover in 1998, a sample of 733,378 individuals had been insured with the same fund in 1997; this sample was retained for the prospective analyses. Partial year insurees in the 1997 and 1998 populations included those who died, the newborn, recent entrants, and those who opted out of the fund; none of them was excluded (see table 1 at the end of the paper).

A claims history file for all hospitalisations ending in 1997 and 1998 respectively provided data on the length of stay, the principal 3- or 4-digit International Classification of Diseases, 9th revision (ICD-9) diagnostic code on discharge, and the charges for each episode of inpatient care and rehabilitation. Applying a US-American diagnosis-based risk adjustment model to German data necessitates a crossover of German ICD-9 to US-American ICD-9, Clinical Modification (ICD-9-CM) codes. This study used a crossover itemised in [1].

The study data files further contained an insuree's annual health care expenditure by type of service, including annual per-person payments for hospital care, ambulatory care (provided by both general practitioners and specialists), prescription drugs, dental care, ancillary services, durable medical equipment, and home health, as well as sickness benefits. The total annual expenditure per insuree was calculated by summing up individual payment amounts accross the different types of service during the calendar year. Spa treatment payments were not included in this total since they are not allowed for in Germany's risk adjustment scheme; neither were sickness benefits, because these benefits in cash are income-related. All

expenditure data of this study is reported after co-payments for those not exempted from cost sharing are deducted.

Except for annual per-person payments for ambulatory and dental care, all socio-demographic, diagnostic, and expenditure information was derived from generally compiled computerized data files. Individual payment data for ambulatory and dental care were a "by-product" of an experimental no-claims bonus arrangement effective in 1997 and 1998 that allowed the refund of a month worth contribution rate if an insured person had had no claims on curative medical services in a calendar year. Providing the sickness fund with expenditure information on individual services was a prerequisite for this arrangement – normally, the physicians' and dentists' self-governing bodies in charge of the reimbursement of office-based physicians and dentists do not release any insuree-level claims data to sickness funds. The data were reviewed for general plausibility and validity. Where appropriate, the sickness fund submitting the data was contacted to verify problems and identify possible solutions.

2.2 Methods

Predictive performance

Risk assessment models for study

The following models were evaluated: a demographic model, the current socio-demographic model, a model incorporating a simple binary indicator for hospitalisation, and a DCG/HCC model.

Demographic model: Age and gender

Preliminary analyses for this basic age/gender model had shown that there is hardly any difference in the explanation of expenditure variance when using broader age clusters instead of 1-year age groups (from 0 to ≥90) as laid down in Germany's current risk adjustment scheme. Clearly 1-year age groups make a weights matrix cumbersome to operationalise. The number of age/gender entries was therefore reduced to 26 age/gender groupings, 13 each for females and males aged 0-5, 6-13, 14-17, 18-25, 26-35, 36-45, 46-55, 56-60, 61-65, 66-70, 71-75, 76-80 and ≥81 respectively.

Socio-demographic model: A binary indicator for invalidity with age and gender

Over and above age and gender, the model differentiates persons who are entitled to an invalidity pension from all others, thus simulating Germany's current socio-demographic risk

assessment methodology. Insurees who did not draw an invalidity pension were assigned to one of the 26 age/gender groups described above. Recipients of an invalidity pension – entitlement to such a pension is possible from age 18 to 65 – were in turn assigned to 14 age/gender groups with age groups for ≤ 35, 36-40, 41-45, 46-50, 51-55, 56-60 and 61-65 years. For the purposes of risk assessment, invalidity status was defined as entitlement to invalidity benefits in any single month during the risk year (i.e., all or part of 1997 in both the concurrent and the prospective framework). Interaction among age/gender and invalidity is accounted for.

Hospitalisation model: A binary indicator for hospitalisation with age, gender, and invalidity status

The model is an extension of the socio-demographic model that tries to further distinguish potential high-expenditure individuals by relatively simple means. Individuals were also categorised based on their utilisation or non-utilisation of inpatient services (hospital inpatient stays of any length) in the risk year. Interactions between age/gender/invalidity and hospitalisation are accounted for.

DCG/HCC model: Hierarchical condition categories (HCCs) with age, gender, and invalidity status

In addition to the age/gender groupings described for the demographic model and the invalidity indicator, the model uses all of the reported inpatient diagnoses to categorise insurees into disease groups according to the DCG/HCC classification methodology. Preliminary analyses showed that accounting for interactions between age/gender and invalidity did not add substantially to the predictive performance of the diagnosis-based risk assessment model; hence the model adopts a simple additive relationship between age/gender, invalidity, and multiple diagnostic categories.

Insurees' diagnostic classification in this study was implemented by $D_xCG^{®}$ Risk Adjustment Software, Release 6.1 (August 2002). This DCG/HCC grouping software version maps the ICD-9-CM diagnosis codes to 781 clinically homogenous groups (DxGroups), which are in turn grouped into 183 clinically related and resource homogenous condition categories (CCs). To exclude that minor diagnoses add to expenditure predictions, CCs are arrayed in hierarchies of related CCs, the HCCs.

No amendments were made to insurees' diagnostic classification obtained from the grouping software save lumping HCC categories with less than eight cases in 1997 to form a residual

risk category with enhanced statistical stability. Diagnostic groupings that contained no patients were discarded.

In addition to a matrix of dichotomous diagnostic categories, the grouping software produces predicted expenditure scores, i.e. relative risk weights that have been normalised to have a weighted mean of 1.0 in the original benchmark sample on which they were developed[3]. The study examined the usefulness of applying this "off-the-shelf" version of the DCG/HCC model to German sickness fund data: Besides a fully reparameterised model, an offered weights model was also calculated with the original, software-computed DCG/HCC relative risk weights (RRWs), the age/gender groupings and the invalidity indicator as insuree classifying variables[4].

Parameter estimation

A series of multivariable linear regressions was constructed to examine the ability of each of the risk assessment models summarised in table 1 in predicting year-1 and year-2 expenditure respectively. Specifically, year-1 (concurrent framework) or year-2 (prospective framework) annual expenditure was regressed on insurees' year-1 classifying variables that define the respective models[5]. All independent variables apart the RRWs produced by the grouping software in the offered weights model were entered as class, rather than continuous variables. The analyses presented in this article focused on the models' predictive performance and did not dwell on the statistical significance of the included classifying variables. The specification of each model was forced; potential negative parameter estimates remained included and were not set to 0.

Several researchers have proposed methods other than single-equation linear regression to estimate health services utilisation and expenditure because the distributional properties of these data are of statistical concern and may require transformation[6]. Traditionally, linear regression has been the technique of choice for predicting medical risk. Research has

---

[3] The offered risk weights were calibrated to total covered charges (fee-for-service and managed care) from the DCG/HCC model benchmarked to the 1997 to 1999 commercial population.

[4] Although the RRWs incorporate age and gender, the socio-demographic factors were entered separately in the regressions in order to recalibrate expenditure predictions, i.e., to have the predicted mean expenditure being equal to the actual mean expenditure in the German concurrent and prospective samples respectively.

[5] The single-equation WLS regressions that were constructed were in the form of $Y_{it} = X_{it}\beta_i + U_i$ and $Y_{it} = X_{it-1}\beta_i + U_i$ in retrospective and prospective modelling respectively, where Y_{it} are annualised health care expenditures for the i^{th} person in year t (t = year 1 in the concurrent framework and t = year 2 in the prospective framework), X_{it} and X_{it-1} are the demographic and diagnostic characteristics for the i^{th} person in year t (t = year 1 in the concurrent framework) or t-1 (t = year 1 in the prospective framework), β_i are the coefficients associated with each of the demographic and diagnostic characteristics and U_i is a disturbance term.

[6] See for example [2]

demonstrated that ordinary least squares (OLS) or weighted least squares (WLS) regression is quite robust to asymmetric and highly skewed errors, though it ignores the mixed character of the underlying distribution of expenditure; with large sample sizes, even adequate efficiency can be achieved. Moreover, using OLS or WLS regression that retains the original scale of the response allows easy and meaningful calculation of an individual's risk profile by summing coefficients for each descriptor. Hence, for the sake of simplicity, robustness, and the ease of direct interpretation, this study stayed with employing the linear regression method and untransformed expenditure (i.e., euros)) rather than using alternative estimation approaches.

Also, to predict per insuree per year (PIPY) expenditure and accommodate partial year insurance coverage in the prediction year – year 1 in the concurrent framework, year 2 in the prospective framework –, a weighting algorithm was used: Total per-person expenditure in the prediction year was annualised by dividing actual expenditure by the fraction of the prediction year that an individual had been insured. In subsequent calculations of means and regressions, each insuree's annualised expenditure was weighted by this same fraction. Annualising and weighting observations is needed to compute unbiased estimates of mean and total expenditure per year when each observation corresponds to a different sample size (in this case, the fraction of the year an individual is covered).

In order to mitigate the potentiality of overfitting and avoid estimates of predictive accuracy that are upwardly biased, a 10-fold cross-validation approach was used: The data were randomly split into ten disjoint sets of nearly equal size which define ten different splits into calibration and validation sets, one-tenth of the data being the 1st,…, 10th validation set and the remainder of the data the 1st,…, 10th calibration set[7]. Calculation required ten model calibrations; for each of the ten splits, measures of predictive accuracy were derived by computing expenditure estimates from the calibration set (within sample estimates) and applying them on the respective validation set as expenditure predictions (out of sample predictions). Predictive accuracy was then computed as the mean of the estimates of predictive accuracy for the ten validation data sets. A separate calibration-validation analysis was performed for each model in the concurrent and prospective applications [3].

[7], Practical experience with K-fold cross-validation suggests that a good strategy is to take $K = \min(n^{1/2},10)$, on the grounds that taking K>10 may be computationally too intensive when the prediction rule is complicated, while taking groups of size at least $n^{1/2}$ should perturb the data sufficiently to give small variance of the estimate.

Measures of predictive performance

Several descriptive measures on the individual and group level were computed to gauge the models' relative predictive performance. Individual level predictive performance was measured using individual adjusted R-squared, the mean absolute prediction error, and Cumming's prediction measure; group level predictive performance was assessed employing predictive ratios of expenditure quintiles.

Individual adjusted R-squared (R^2). R^2, the conventional regression-computed measure used to estimate model fit, describes the proportion of the individual variance in actual expenditure that is explained by a model. To use the measure for estimating predictive accuracy, R^2 values were produced over all observations in the validation data through applying the formula

$$R^2 = 1 - ([\sum_i (a_i - \hat{a}_i)^2] / [\sum_i (a_i - \bar{a})^2])$$

where a_i is actual year-1 or year-2 expenditure for person i, $\hat{a}_i$ predicted year-1 or year-2 expenditure for person i, and $\bar{a}$ the mean of actual year-1 or year-2 expenditure (i goes from 1 to n, where n is the number of observations). In order to compare the predictive performance of models that varied substantially in complexity, the study computed adjusted R^2 values[8].

Mean Absolute Prediction Error (MAPE). MAPE is defined as the mean of the absolute difference between actual and predicted expenditures across all individuals [4]. With this type of measure, predictions that are greater or less than actual expenditure cannot cancel each other out, as can happen with the mean prediction error. MAPE values over all observations in the validation data were derived as

$$MAPE = (\sum_i |a_i - \hat{a}_i|) / n$$

Cumming's Prediction Measure (CPM). The concept of using the absolute value of the prediction errors rather than the square of the prediction errors as with R^2 when trying to summarise the predictive performance of various models on an absolute basis was developed in Cumming et al. [4]. It arose from concern about the sensitivity of R^2 to large prediction errors. CPM values over all observations in the validation data were calculated as

$$CPM = 1 - ([(\sum_i |a_i - \hat{a}_i|) / n] / [(\sum_i |a_i - \bar{a}|) / n]) = 1 - ([\sum_i |a_i - \hat{a}_i|)] / [\sum_i |a_i - \bar{a}|])$$

[8] Adjusted R^2 allows for the degrees of freedom of the sums of squares associated with R^2. Therefore, even though the residual sum of squares decreases or remains the same as new independent variables are added, the residual variance does not. Unlike R^2, adjusted R^2 can decline in value if the contribution to the explained variance by an additional variable is less than the impact on the degrees of freedom. Adjusted R^2 is calculated as $R^2_{adj.} = 1 - ([1 - R^2] * [(n - i) / (n - p)])$ where n is the number of observations, and p the number of independent variables including the intercept, with i = 1 if there is an intercept and i = 0 otherwise.

Predictive ratio (PR) of expenditure quintiles. The PR is a group measure and can be calculated as the ratio or the aggregate predicted year-1 or year-2 expenditure for a given group of insurees g divided by the aggregate actual year-1 or year-2 expenditure for the same group g:

$$PR_g = \sum_{ig} \hat{a}_{ig} / \sum_i a_{ig}$$

The comparison gives the reciprocal of the common observed-to-expected actuarial ratio. A model predicts well for a group of insurees when it's PR is close to 1.0; a PR greater than 1.0 indicates overprediction, whereas a PR less than 1.0 indicates underprediction. Risk assessment models may generate predictions of differing accuracy for various ranges of the expenditure distribution; the study thus calculated PRs for groups of insurees defined by quintiles of (non-annualised) actual expenditure, assessing the strength of each model to predict expenditure for relatively high, medium, and low expenditure subjects.

Additional investigations

Truncation of expenditure

To study the application of outlier risk-sharing in the context of risk adjustment, the models were also calculated using truncated expenditure. A first approach simulated the effect of an outlier threshold of € 20,000 in *actual* non-annualised expenditure on the predictive performance of the models; cases that exceeded the threshold were top-coded at the capped amount[9]. The second approach fixed a € 10,000 deductible above the non-annualised expenditure *predicted* by the models, thus simulating the effect of a variable outlier threshold on the predictive performance of the models.

Leavers versus Joiners versus Stayers

Partial year insurees, i.e., "leavers" – those with any insurance cover in year 1 but not in year 2 – and "joiners" – those with any insurance cover in year 2 but not in year 1 – present specific concerns with respect to risk assessment and adjustment. The main concern is that those individuals among the leavers and joiners who switch sickness funds (i.e., entrants and individuals opting out) may represent unique socio-demographic and health statuses relative to insurees with ongoing cover (stayers) and that they are, on average, much healthier and less expensive than the stayers, which lays the basis for potentially severe biased selection. Also,

[9] The capped amount approximates the annual threshold of € 20.450 for the outlier risk-sharing arrangement of German SHI. German sickness funds would pay another 40% of the expenditure above the threshold amount and are compensated for the remaining 60% of the expenditure above the threshold out of the risk pool; this was not simulated in this study.

with a prospective approach, analyses can only be performed using data for individuals with information about insurance cover for at least two years – diagnostic and/or demographic information of year-1 and year-2 length of cover and expenditure data) –, so analysis may have a selection bias.

This study evaluated the concurrent findings for both study years, 1997 and 1998, separately for different sub-groups that composed the leavers, joiners, and stayers. Among the leavers, those who died in 1997 were distinguished from those who opted out of the fund (i.e., the switchers among the leavers); among the joiners, the newborn of 1998 were distinguished from entrants still alive at the end of 1998 (i.e., the switchers among the joiners); and among the stayers, distinction was drawn between the newborn of 1997, older insurees who were still alive at the end of 1998 (i.e., the non-switchers), and those who died in 1998. To assess the issue of turnover in the sickness fund's population, the concurrent mean expenditure predicted under each risk assessment model was compared with the actual mean expenditure separately for the switchers among the leavers, the switchers among the joiners, and the non-switchers. To explore the sensitivity of the prospective findings to the select group of insurees who had ongoing cover in both study years, model-specific concurrent predictive ratios for the newborn and the decedents among the leavers, joiners, and stayers were computed separately and tested for systematic over- or underprediction.

Entitlement to LTC insurance benefits

An additional analysis was performed to test the explanatory power of a binary indicator for persons who were entitled to benefits of long term care insurance (LTC). For that analysis the concurrent and prospective analyses were repeated using this indicator in lieu of and in interaction with the invalidity status variable. For the purposes of risk assessment, LTC status was defined as entitlement to LTC insurance benefits in at least one month during the risk year (i.e., all or part of 1997 in both the concurrent and the prospective framework).

All additional investigations were performed using the overall concurrent and prospective samples without reserving portions for validation. With the large sample sizes used, the ten validation data sets yielded on average values for the performance measures that were similar to those found when estimating the models on the calibration data sets or on the overall samples. The DCG/HCC model using offered weights was no longer used, since it was found to be inferior in predictive performance to the reparameterised DCG/HCC model.

SAS (version 8.2) was used for all analyses.

2.3 Literature in Section "Data and Methods":

[1] Deckenbach, B., et al., *Diagnose-basierte Risikoschätzmodelle für die Patientenklassifikation im vertragsärztlichen Vergütungssystem in Deutschland. Überprüfung der empirischen Anwendungsvoraussetzungen und exemplarische Analyse der Gruppierungsvorgänge zur Beurteilung der Übertragbarkeit von ausgewählten medizinischen Klassifikationssysteme. Gutachten für die Kassenärztliche Bundesvereinigung.* 2001, IGES: Berlin.

[2] Jones, A., *Health Econometrics*, in *Handbook of Health Economics*, A.J. Culyer and J.P. Newhouse, Editors. 2000, Elsevier: Amsterdam. p. 265-344.

[3] Davison, A. and D. Hinkley, *Bootstrap Methods and Their Application*. 1997, Cambdridge: Cambridge University Press.

[4] Cumming, R.B.K., David; Cameron, Brian; Derrick, Brian, *A Comparative Analysis of Claims-based Methods of Health Risk Assessment for Commercial Populations*. 2002, Park Nicollet Institute Health Research Center: Minneapolis. p. 62.

3. Results

3.1 Descriptive statistics

Descriptive statistics for the retrospective and prospective study samples are provided in Table 1. The sickness fund population was about 52% female in both years, with a mean age of approximately 46 years (SD 23.7); around 25% of insurees (33% of women, 16% of men) were age 66 and older. Most insurees (87.6%) had 12 months of cover in any study year. The mortality rate was 1.8% in both years.

Annual health care expenditure in these samples, as elsewhere, was highly skewed. Year-1 and year-2 expenditure average € 1,740 and € 1,762 respectively, with standard deviations being roughly 3 times larger than the mean. Only a small fraction of the population (about 5.5%[10]) did have no encounter with the health care system in 1997 or 1998 respectively. Also, in the prospective sample, year-2 expenditure of those who incurred no health care expenditure in year 1 was substantially less than (i.e., 32% of) average. The 1% of insurees whose expenditure was highest in any year used up about 20% of that year's total health care resources. For these high-users who had per-person expenditure of € 20,000 and more, truncated expenditure represented a little less than 60% of their total resource use. Roughly two-third of health care resources were consumed by 10% of insurees; reciprocally, the about 50% of insurees with individual expenditure up to € 500 used less than 6% of total resources. In the prospective sample, the 1% of insurees with the highest year-1 expenditure used up about 9% of year-2 total health care resources.

Between 16% and 17% of insurees were hospitalised in any study year, with inpatient expenditure comprising the largest portion (45-47%) of total reported expenditure. Not shown in table 1 is the clinical characterisation of the study population as implemented by the DCG/HCC grouping software. A total of 210,574 patient/diagnosis pairs for 127,905 (16.8%) insurees ever hospitalised in 1997[11] were submitted to the grouping software, corresponding to an average of 1.4 (SD 0.8, range 1-17) unique diagnosis codes per person hospitalised. About 38% of the hospitalised insurees were age 66 and older. Seventy-seven of the 210,574 diagnoses submitted were incompatible with age or gender, 579 were identified as

[10] Though substantially less than the percentage reported for the US (about 25%) or the Netherlands, this figure is in accordance with the findings by Breyer et al. (2003) for a sample of insurees covered by several smaller German sickness funds.

[11] Over 90% of insurees (i.e., 115,150) hospitalised in 1997 (year 1) were also covered by the fund in 1998 (year 2).

numerically invalid (549 of these invalid entries featured the non-existent code 613). A total of 32.5% of all diagnosis codes were 3-digit; V codes (factors influencing health status) comprised less than 0.05% of all codes, and there were no E codes (external causes of injury and poisoning). All insurees ever hospitalised were assigned to at least one HCC, if they hold a valid diagnosis code. Hospitalisations were primarily attributable to heart conditions (Aggregated Condition Category (ACC) 16 with 19.161 patients, i.e. a rate of 243/10.000), gastrointestinal conditions (ACC 07: 14.828 patients, rate 188/10.000), injury, poisoning and other complications (ACC 26: 14.550 patients, rate 185/10.000), malignant neoplasms (ACC 02: 10.054, rate 128/10.000), and musculoskeletal and connective tissue conditions (ACC 08: 9.900 patients, rate 126/10.000). The most frequently occurring HCCs were HCC 036 ("Other Gastrointestinal Disorders" with 7.482 cases), HCC 127 ("Other Ear, Nose, Throat, and Mouth Disorders": 7.400 cases), HCC 043 ("Other Musculoskeletal and Connective Tissue Disorders": 6.693 cases), HCC 162 ("Other Injuries": 5.283 cases), and HCC 084 ("Coronary Atherosclerosis/Other Chronic Ischemic Heart Disease": 4.992 cases). A number of HCC categories occurred rarely in the dataset. Eighteen HCC categories had no observation in 1997; a further 6 of the 165 categories with observations had frequencies less than 8. The *unadjusted* and mean year-1 expenditure of the HCC categories exhibited a roughly 18-fold variation, ranging in round terms from € 2,000 for HCC 142 ("Miscarriage/Early Termination") to € 36,000 for HCC 131 ("Renal failure"), which underlies the importance of risk adjustment. However, many HCC categories, especially those with low frequency, had quite wide 95% confidence intervals; two categories included negative values in their confidence intervals, for four categories with a frequency of 1 each no confidence interval was computable.

3.2 Predictive performance

Individual level predictive performance

Table 2 summarises the predictive performance of the four risk assessment models and the DCG/HCC model with offered weights as measured by the R^2, MAPE, and CPM statistics. Both concurrent and prospective findings are reported; all of the presented statistics are means of the estimates of predictive performance computed for the ten validation data sets in 10-fold

cross-validation[12]. Higher R^2 and CPM values and lower MAPE values indicate better predictive performance.

Concurrent approach. As expected, the demographic age/gender and socio-demographic age/gender/invalidity models were the least predictive models, regardless of the measure used. There were substantial increases in predictive performance, however, when insurees who were hospitalised were distinguished from those who were not, and even more when the diagnostic information for those hospitalised was taken into account. Based on adjusted R^2, the demographic model explained little more than 3% of the variance in total actual expenditure[13]. Incorporating invalidity status brought a relatively modest gain with an adjusted R^2 of 5.1 %. The simple addition of a binary variable for hospitalisation achieved an adjusted R^2 value of about 25%, nearly an 8-fold improvement over the demographic model. Finally, with an adjusted R^2 value of more than 37%, the predictive performance of the reparameterised DCG/HCC model represented a 12-fold improvement over that of the demographic model.

MAPE and CPM provided the same rankings of the models' predictive performance as adjusted R^2, though the magnitudes of the models' relative improvements differed across the measures. By decreasing MAPE from about € 1,850 to around € 1,200 and € 1,000 respectively, the hospitalisation and reparameterised DCG/HCC models outperformed the demographic model by 33% and 42%. With CPM values of roughly 38% and 46%, the two models performed roughly 5.5 times (hospitalisation model) and 7 times (reparameterised DCG/HCC model) better than the demographic model (CPM of 7%) did.

Table 2 also shows the significant increase in performance due to reparameterising the DCG/HCC model to the German sickness fund population rather than using the HCCs with the offered relative risk weights. This might be expected since the standard set of relative risk weights was calibrated for a US commercial population. Reparameterisation of the DCG/HCC model increased its general performance by approximately 20% as measured by adjusted R^2, by 9% based on MAPE, and by 13% when using CPM as the measure.

---

[12] When averaged over the ten validation sets, the estimates of predictive performance computed in 10-fold cross-validation differed only insignificantly from the averaged estimates obtained in the calibration data sets or when fitting the models on the entire concurrent and prospective samples, indicating a good *overall* fit between predicted and actual expenditure for all models. Averaged year-1 and year-2 predicted means were equal to actual year-1 and year-2 mean expenditure.

[13] The reported concurrent and prospective R^2 values for the demographic model, though small, are higher than those found in most of the international studies. This might be due to a comparatively lower expenditure variance in German data: while US and Dutch studies display coefficients of variation of more than 4.0, the coefficients of variation determined in the present study were below 3.0 (and thus in accordance with the one reported by [1]).

The mean expenditure for recipients of invalidity benefits was substantially higher than predicted by age/gender and HCCs alone. Although incorporating invalidity status in the DCG/HCC model had only a relatively small impact on individual performance measures – for instance an increase of adjusted R^2 from 36.9% for a reparameterised age/gender-HCCs model (results not shown) to 37.3% for the reparameterised age/gender-invalidity-HCCs model –, the analyses showed that, if this factor was not taken into account in the estimations, the model would not predict the average expenditure of this important and easily observed higher-expenditure subgroup, leading to an underpayment of these insurees if such an adjustment was not included.

Prospective approach. The prospective findings in table 1 show the same rankings of the models' predictive performance as the concurrent results. However, when changing from the concurrent to the prospective application, the health-status based risk assessment models predicted much less of the variation in year-2 expenditure than the concurrent models did for year-1, whereas the difference is almost negligible for the demographic and demographic/invalidity model; therefore the gain in predictive performance for these models over the demographic model was thus much smaller. The prospective R^2 values ranged from 3.1% for the age and gender model to 7.1% for the hospitalisation and 11.7% for the reparameterised DCG/HCC models. When using MAPE as the measure, the decrease from € 1,902 to € 1,786 and € 1,738 corresponded to a gain in predictive performance for the hospitalisation and the reparameterised DCG/HCC models over the demographic model of roughly 6% and 8.5% respectively. Based on CPM, the difference in the relative improvement between the hospitalisation and the reparameterised DCG/HCC models was even less pronounced, the hospitalisation model (CPM of 13.2%) performing 2.1 times and the DCG/HCC model (CPM of 14.4%) 2.3 times better than the demographic model (CPM of 6.3%).

Group level predictive performance

The relative rankings of the models remained unchanged when their concurrent and prospective predictive performances for groups of insurees with relatively high, medium, and low expenditure were evaluated. Table 3 displays the PR results for insurees grouped by quintiles of (non-annualised) actual expenditure. Quintile 1 (Q-1) represents the 20% of the population that had the lowest expenditure and quintile 5 (Q-5) represents the 20% of the population that had the highest expenditure. Specifically, Q-1 had actual year-1 and year-2 expenditure of € 53 and € 55 per insuree per year (PIPY) respectively and Q-5 of € 6,584 and

€ 6,812 PIPY respectively (or about 3.8 times the overall year-1 and year-2 expenditure mean of € 1,740 and € 1,762 respectively).

Not surprisingly, the age and gender model is grossly overpaying for the low expenditure quintile Q1, giving a PR of 24 (concurrent) and 23 (prospecitve) for this group of insured; on the other hand it is grossly underpaying for the high expenditure quintile Q5, giving a PR of 0.3 (both, concurrent and prospecitve). In comparison to that, all other models perform better. In line with the individual level results, the reparameterised DCG/HCC model did perform best in both the concurrent and prospective application and over all expenditure quintiles; however also with this model there is still over-payment for the low-expenditure insured (with a PR of 21 concurrent and prospective) and under-payment for the high-expenditure insured (with a PR of 0.8 in concurrent and 0,5 in the prospecive application). The PR values obtained with the simple binary hospitalisation variable, almost reached the results for the reparameterised DCG/HCC model.

In all modells, the prediction errors were larger for insurees with expenditure levels farther from the mean. With a prospective model application, the differences between the models' predictive performances are small for the middle 60% of the expenditure distribution; the greatest difference in these middle three expenditure quintiles amounts to a 22% more accurate (over)prediction for the reparameterised DCG/HCC model (PR of 4.8) relative to the demographic model (PR of 5.9). In concurrent application, the differences between the models' predictive performances in the middle 60% of the expenditure distribution are greater; the smallest difference was observed in the 4th quintile with the DCG/HCC model (PR of 1.1) giving a 63% better (over)prediction of group expenditure than the demographic model (PR of 1.8).

For both the least and most expensive insurees, all four models did a more or less poor job in predicting expenditure concurrently or prospectively; similar to the findings for the middle three expenditure quintiles, the differences between the models' predictive performances in the two outer quintiles were much more pronounced in the concurrent than in the prospective application. The PRs in these two quintiles reflect the range and degree of “skewness” of expenditure predictions. PRs for Q-1 were quite high mainly because actual expenditure levels in this quintile were as low as or closely above zero whereas none of the risk assessment models could predict an expenditure close to zero. In their concurrent and prospective applications, the minimum predicted annual amounts ranged from € 566 and €

649 for the demographic model to € 183 and € 486[14] for the reparameterised DCG/HCC model respectively. At the other end of the expenditure distribution, PRs for Q-5 were below 1.0 as no insuree could be expected to have expenditure levels 300 times above average. The maximum annual prediction values for the reparameterised DCG/HCC model totalled to € 72,623 when applied concurrently and € 58,458 when applied prospectively, i.e., not even a 50-fold of the mean. The reparameterised DCG/HCC model thus achieved an extended upper tail by predicting less in the lower and middle parts of the expenditure distribution. In contrast, the demographic model's maximum predicted annual expenditure – € 3,732 for the concurrent and € 3,712 for the prospective application – amounted to a little more than twice the mean.

3.3 Additional investigations

Truncation of expenditure

To simulate the effects of outlier risk-sharing on the models' predictive performance, two different approaches of truncating actual expenditure were applied. The results in table 4 are based on the truncation of expenditure at a pre-specified threshold, i.e., cases exceeding € 20,000 in actual, non-annualised expenditure were top-coded at that amount. The results in table 5 pertain on the other hand to the application of a variable truncation threshold in that actual expenditure was truncated at a € 10,000 deductible above the non-annualised predicted expenditure.

None of the truncation approaches appeared to cause any significant changes in the overall relative performance of risk assessment models compared with each other. In general, for both the concurrent and prospective applications, similar model-specific improvements in each of the measures were observed: expenditure truncation induced an increase in the adjusted R^2 and CPM values, a decrease in the MAPE values, and tended to bring the PRs closer to 1.0. The improvements were not dramatic, though, with MAPE, CPM, and the PRs being less sensitive to expenditure truncation than adjusted R^2. The CPM measure for instance showed a model-specific increase by 5-10% in the pre-specified truncation threshold approach and by 10-20% in the variable truncation threshold approach as compared to the findings of table 2 based on untruncated expenditure; in terms of adjusted R^2, the model-specific increases were

[14] The prospective reparameterised DCG/HCC model also generated negative predicted expenditure for three individuals aged 18 and younger who had all three hospitalised in year 1 (and in year 2) with a diagnosis included in HCC 5 ("Opportunistic infections").

around 40-90% and 70-160% respectively. Given that R^2 squares the prediction errors, it can be overly affected by a relatively small number of high-expenditure case and increases with lower thresholds for truncation; this is less an issue with MAPE and CPM. The more health status information the models included, the less the measures of predictive performance improved when expenditure was truncated. The differing mean expenditure levels in the variable truncation threshold approach suggested that the better a risk assessement model predicts the upper tail of the expenditure distribution and reproduces the skewness of health care expenditure, the less important in magnitude the reallocations of the outlier risk-sharing approach will be.

Leavers versus Joiners versus Stayers

Approximately 7% of the individuals insured with the sickness fund in 1997 were not in the fund in 1998 (leavers). Of those covered by the fund in 1998, 3% were not in the fund in 1997 (joiners). The mean expenditure for the leavers was roughly 2.5 times as high as that for insurees who were in the fund in both study years (stayers), whereas the mean expenditure for the joiners was about 15% lower as compared to the mean expenditure of stayers.

The three groups comprise varying percentages of newborn, decedents, and/or individuals entering or opting out of the fund. The leavers' relatively high mean expenditure was primarily due to the high expenditure of insurees in their last year of life. With € 14,800 PIPY, expenditure for leavers who died in 1997 was similar to that for stayers who died in 1998. Also, the switchers among the leavers and joiners – i.e., those opting out (leavers who were still alive at the end of 1997) and those entering (joiners who were still alive at the end of 1998) – were, on average, younger, healthier, and only half as expensive as the non-switchers (stayers who survived throughout 1997 and 1998) – the differences in expenditure ranged from € 657 PIPY for the leaving switchers to € 760 for the joining switchers.

Figure 1 depicts the differences between predicted mean expenditure and actual mean expenditure under each risk assessment model for the three sub-groups of the non-switchers, the switchers among the leavers (switchers: out), and the switchers among the joiners (switchers: in). A positive difference involved a predictable gain for the sickness fund; a negative difference equalled a predictable loss. In a concurrent risk adjustment scheme based on the demographic model, switchers represented "good risks" insofar as their mean predicted expenditure was higher than their actual mean expenditure and even higher than the mean predicted expenditure for the non-switchers. Using a socio-demographic risk assessment model reduced the demographic model's mean predictable gain of 1997-€ 198 for the leaving switchers and 1998-€ 262 for the joining switchers by roughly 60%. The hospitalisation

model reduced the potential gains for leaving and joining switchers substantially, converting these individuals, on average, to "bad risks"; the mean predictable loss ranged from 1997-€ 74 to 1998-€ 8 respectively (compared to the demographic model's mean predictions, the loss amounted to roughly € 270 for both for the leaving and the joining switchers). The DCG/HCC model further increased the predictable loss to 1997-€ 87 PIPY for the leaving switchers and to 1998-€ 35 PIPY for the joining switchers. As for the non-switchers, the more health information the risk assessment model incorporated, the better the mean predictions matched the actual mean expenditure for this sub-group.

The results summarised in Table 6 show that the model-specific predictive ratios for the two sets of comparison groups – the decedents, i.e., leavers dying in 1997 and stayers dying in 1998, and the newborn, i.e., joiners born in 1998 and stayers born in 1997 – were similar. For the newborn, no systematic differences in the ratios across models were observed, the ratios being all close to 1.0. In contrast, the predictive ratios for decedents varied systematically across models: the DCG/HCC model performed roughly 3.4 times better than the demographic model, significantly reducing the underpredictions for this group. The decrease of the underpredictions for decedents ran concurrent to the reduction of the model-specific overpredictions for the sub-group of the non-switchers.

Entitlement to LTC insurance benefits

The findings in tables 2 and 3 suggest that the invalidity indicator may be considered as a weak proxy measure for the unique expenditure implications of characteristics not related to hospital admissions. The results in table 7 indicate that the same holds for the LTC indicator. Using it in lieu of the invalidity indicator showed model-specific concurrent and prospective predictive performances similar to that of the invalidity indicator's one. Using the LTC indicator in combination with the invalidity indicator further improved, albeit marginally, the individual and group level predictive performances for all models with the exception of the hospital model. When applied concurrently, the latter model performed less well in predicting individual expenditure when incorporating both, the invalidity indicator and the LTC indicator, than when using either of them; the concurrent group level predictive performance in contrast improved considerably with the inclusion of both indicators.

3.4 Literature Section “Results”

[1] Breyer, F., *Determinants of health care utilization by German sickness fund members-with application to risk adjustment.* Health Economics, 2002(10/2002).

4. Discussion

4.1 Discussion of findings

Effective risk adjustment is a crucial aspect of competitive health insurance system with non-risk related premiums. The effectiveness of any risk adjustment approach will depend on accurate risk assessment. Germany's Risk Equalisation Reform Act of 2001 makes provisions for implementing a risk assessment formula that accounts for insurees' health status, as this may help reorient the current incentive structure in the sickness fund market and reduce the negative consequences of enrolling high-risk users by compensating sickness funds for health care needs which persits with the existing socio-demographic risk adjustment formula.

The 2001 legislation explicitly states that pertinent international risk assessment methods be adopted. The limited availability, completeness and validity of administrative health information to German sickness funds, however, potentially threaten the applicability and validity of these methods. Moreover, their external validity may be questioned as the underlying disease classifications are based on data of populations with specific demographics, coverage, utilisation and provider practice patterns, coding methodologies, provider reimbursement policies, and prices.

The issue of applying and generalising such methods to the German situation thus necessitates thorough investigation. This is the first published study that explores the use of a US-American diagnosis-based risk assessment approach on data of a German sickness fund population and compares its performance to that of models with socio-demographic variables alone or in combination with an indicator for use of inpatient services as indirect markers for health status.

The investigated models are based on the most prevalent and accessible source of health information in Germany, claims for inpatient utilisation. Applying the DCG/HCC classification system to the available data proves to be technically feasible. The analyses of the HCC-specific expenditure levels suggest that the HCC system applied to German data shows substantial face validity in quantifying morbidity on the basis of expected levels of health care expenditure for sickness funds. Despite the sole use of 3- and 4-digit ICD-9(-CM) codes, the HCC system performed reasonably well. However, the system could perform differently (and presumably better) if coding was enhanced – some DxGroups and consequently HCCs are, for instance, not created when applied to ICD-9 data because of missing fifth digits (e.g., (H)CC 20, type I diabetes mellitus, requires the specificity of the

five-digit ICD-9(-CM) code to indicate the type of diabetes mellitus). The descriptive inpatient-only data of this study's sample show that DxCG's clustering of DxGroups in condition categories (CCs) on the basis of medically related problems with similar expected expenditure is not always indicated with the mean expense for DxGroups. Also, the descriptive data show in part large variances in estimated mean expenditure for HCCs, especially in groupings with low frequencies. Groupings with low frequencies may represent unstable values for estimation purposes. This matter is vitally important for implementation of risk adjustment, as capitation payments for cases with low prevalence may be incorrectly estimated with consequential over- or under-funding. Even with a larger sample then that available for this study, it might be necessary to pool HCCs (or even DxGroups) to diagnostic subgroups of a specified minimum sample size to obtain parameter estimates that are statistically robust.

The results of this study indicate that, at the individual level, each of the utilisation-based risk assessment models is able to explain a much greater proportion of the variance in total health expenditures than the socio-demographic model currently in use. The standard reference model based on age and gender explains less than 4% of the variance in expenditures whereas the reparameterised HCC model could explain around 37% of the variance in concurrent resource use and 11-12% in prospective resource use. The reduction in explanatory power of the prospective models compared with the concurrent ones is expected given that only a portion of future costs is predictable on the basis of past utilisation (or morbidity) patterns. Van de Ven & Ellis argue that around 20% is "the lower bound on the upper bound" of the proportion of variance in expenditures that is potentially predictable in prospective risk modelling.[1] If this is the case, the DCG/HCC model we tested explained about one half of what could be maximally predicted. The concurrent models tend to improve the non-demographic model's explanatory power. As they approach cost-based reimbursement and return, in essence, risk to the consumer, they muddle the incentives to sickness funds for efficiency. For the privately-insured benchmark US-population on which the reference ("off-the-shelf") risk weights were computed, R^2s obtained by the DCG/HCC explanation models on the basis of inpatient diagnoses are 44.6% and 7.9% when used to predict same year and subsequent year resource use respectively. The R^2s obtained in the risk-adjusted models of this study, which could only predicate on the single primary diagnosis per hospital episode, are somewhat lower in the concurrent case and higher in the prospective case. The comparison is rendered difficult by the fact that DxCG's commercial benchmark population is for the most part an under-65 population with a different distribution of illness burden. The

DCG/HCC system has also been independently evaluated in a number of studies that use data from populations other than the original development populations. Duckett & Agius tested the ability of the DCG/HCC model (release 5.2, May 2001) to predict resource use of participants with any use of services in the first-round Co-ordinated Care Trial (CCTs). Based on inpatient diagnostic information, age, gender, socio-economic status and HCCs explained around 46% of variation in concurrent year log-transformed expenditures [2]. With prospective (subsequent) year modelling, explanatory power was weaker, explaining about 18% of variation in total log-transformed expenditures. Based on both inpatient and ambulatory diagnostic information, about 45% and 23% of variation in concurrent and subsequent year log-transformed expenditures respectively were explained. Especially the use of the logarithm of expenditures complicates the comparison of the Australian results with ours. According to Van de Ven & Ellis, using a log transformation inflates the conventional R^2 by about 100%. Also, sample sizes were relatively small and predictions were only made for a population with any services utilisation. Rosen et al. examined the feasibility of adapting the DCG/HCC model (release 4.2) to the population covered by the US Department of Veterans Affairs (VA) who had some health service use during a 12-month sample period [3]. The dependent variables in this study were utilisation-based rather than expenditures. Predictive power for concurrent "off-the-shelf" prediction of ambulatory provider encounters was 7.7%; a combined inpatient and ambulatory visit measure was higher, at 19.4% of variance explained. Reparameterisation of the DCG/HCC model enhanced performance considerably, to 24.4% for predicting ambulatory provider encounters and 31.4% for service days' prediction. Overall, given the social, epidemiological and economic differences in the populations studied as well as the distinctions in the organisation and financing of their respective health care systems, the international comparability and generalisability of predictive power across studies is limited.

4.2 Limitations

In this study, only the principal inpatient diagnosis for each hospital stay was available to estimate total sickness fund's expenditure for its insured. Although inpatient diagnoses-based risk assessment represents a substantial improvement over socio-demographic assessment alone, many insurees – also particularly ill and high-expenditure subsets of insurees – are not hospitalised in a given year, so that their condition histories remain unknown and are not used for upwards adjustment in the inpatient diagnosis-based model. We found that only 16% percent of our sample had had at least one inpatient stay in the base year, but that 95% in total

had encountered the health care system at some time. Thus, in the absence of ambulatory information – for instance diagnoses from outpatient provider encounters or anatomical therapeutic chemical (ATC) codes for prescribed drugs –, the number of persons able to be classified to discriminating risk adjustment categories is substantially reduced. An additional outpatient grouper should undoubtedly improve measures of enrolee health risk.

The invalidity status indicator may be considered as a weak proxy measure for the unique cost implications of characteristics not related to hospital admissions. It indisputably increases the accuracy of the capitation estimates for a higher-than-average risk subgroup of the population; we reason, however, that its predictive power is impaired in that entitlement to disability benefits is restricted to the under-65 and there are no references whatsoever sickness funds' administrative databases to identify the originally disabled aged 66 and older. In this regard, a tantamount, if not more effective alternative to reflect the costs of characteristics not related to hospital episodes is the inclusion of a LTC status indicator in lieu of or in addition to the invalidity status variable (data not shown), as entitlement to LTC benefits is independent of age.

Risk sharing arrangements – such as proportional, outlier, high-risk and condition-specific risk sharing – are considered necessary when risk adjustment is imperfect. The use of such arrangements presents an efficiency-risk selection trade-off. The need for risk sharing protection remains after the implementation of health-based risk-adjustment. However, the structure of this protection may need to be revised. With outlier risk sharing for instance there may be some diagnoses for which the risk-adjusted payment rate may exceed the threshold amount (i.e. the sickness fund's deductible level). Given the possibility of payment rates for particular diagnoses that are above the sickness funds' deductible levels, appropriate modifications to risk sharing arrangements should be made.

Beyond the inclusion of only inpatient diagnostic information, this study has some other limitations that should be considered when interpreting its results. First, there may be limits to the generalisability of our findings. Though our study sample comprised a diverse group of non-elderly and elderly SHI individuals, we did not have data from other sickness fund populations. Second, we used a version of the DxCG with ICD-9-CM codes and had to rely on crossovers between the German ICD-9/10 version and the US coding version. Also, the lack of coded secondary diagnoses and the use of less specific 3-digit ICD-9 codes may bias in favour of lower-acuity HCC assignments. Third, the investigation of the complex practical issues that partial-year enrolees represent was beyond the scope of this study. Partial-year

enrolment can occur when enrolees join or disenroll from a sickness fund and also because of births and deaths. Partial-year enrolment can also result from changes in eligibility that occur throughout the year. For example enrolees can gain, lose, and regain eligibility for exemption from co-payments during the course of a year. Limiting in the prospective analyses the sample frame to enrolees eligible during all or part of 1997 and enrolled during part or all of 1998 means that babies born in the prediction year were not included in the analysis because they do not have any diagnostic experience in the base year. Infants with severe birth defects and neonates who were born prematurely can be substantially more expensive to care for than infants without such problems. Correction by incorporating newborns' expenditures with those of their mothers was not possible, because the anonymous individual identifiers did not allow of linking mothers' and children's data. Excluding partial-year enrolees from the prospective sample frame represents a limitation of this study, because a fair and equitable capitation method must adjust for these enrolees.

Large prediction errors can end up dominating the calculation of R^2. As a result, significant improvements in the predictive accuracy for people with small or medium size expenditure might have little or no impact on the R^2 measure. The corollary to this statement would be that R^2 is unduly insensitive to improvement in predictions for small or medium size expenditure. In calibrating the models in this study, a linear regression model was used which minimizes the means square prediction error and why some researcher argue that R^2 is the most appropriate measure. Accordingly, the R^2 measure corresponds to the way the risk weights are calibrated. The authors of this article agree with Cumming et al., that one should first define what is believed to be the most appropriate measure (or measures) of predictive accuracy and let that drive the way the model calibrated rather than vice versa. If CPM is adopted as a new standard in measuring predictive accuracy, this might impact the way models are calibrated. In particular, calibration methods that attempt to minimise the mean absolute prediction error, rather than mean square prediction error, might lead to further improvements in model performance. It might also be surmised that methods that try to minimize the mean absolute prediction error might lead to more stable and reasonable risk weights since such methods are not impacted as much by a few large claims. The sensitivity of R^2 to the level of expenditure truncation also leads to a variety of opinions regarding what is the "right" or "optimal" level of expenditure truncation for analysing predictive performance. CPM is similar in magnitude to R^2. However, CPM tends to me more stable as the level of expenditure truncation is changed. CPM is sometimes higher and sometime lower than R^2. Use alternative methods of model calibration. Specifically, the impact of using

methods that try to minimise the mean absolute prediction error as opposed to methods that minimise the mean square prediction error.

In conclusion, risk adjustment based on socio-demographic factors is inadequate in predicting resource use, so that health status risk assessment and adjustment should continue to play an important role in Germany's health care reform strategies. The primary shortcoming of an inpatient diagnosis-based model as the one explored in this study is its failure to account for conditions treated in ambulatory settings that predict expense. This shortcoming results in systematic over-payments for healthy enrolees and under-payments for enrolees with serious conditions who were not hospitalised in the same or previous year. A multiple-site model requires much more data than does an inpatient diagnosis-based model, raising concerns about coding practice, data collection and processing, incentive structures, implementation, and cost that will challenge German sickness funds, providers, scientists, and politicians. Limitations associated with data availability and validity suggest the need for further research to develop and test utilisation-based risk adjusters for Germany. One approach might be to construct a combined model that incorporates risk-adjusters form different international methods. However, further research is needed to confirm the appropriateness of this method and to identify which risk-adjusters demonstrate consistency across various subgroups of the German SHI population. Overall, the practical application of a health-based risk capitation payment model will involve "a series of trade-offs between risk adjustment theory, data availability, and the willingness of participants [and stakeholders] to play the game".

Apart from improving risk adjustment it should be emphasized finally, that good risk adjustment is a necessary but not a sufficient condition for a competitive system of health insurance with non-risk related premiums to lead to improvements in efficiency in health care production. What is needed urgently as well is a regulatory framework which allows sickness funds to behave as prudent buyers of health care services – such a framework is far from being implemented in the corporatist structures of the German health insurance system.

4.3 Literature in Section Discussion:

[1] van de Ven, W.P.M.M. and R. Ellis, *Risk Adjustment in competitive health plan markets*, in *Handbook of Health Economics*, A.J. Culyer and J.P. Newhouse, Editors. 2000, Elsevier North Holland: Amsterdam. p. 755-845.

[2] Duckett, S.J. and P. Agius, *Performance of Diagnosis-Based Risk Adjustment Measures in an Population of Sick Australians.* Australian and New Zealand Journal of Public Health, 2002. **26**(6): p. 500-507.

[3] Rosen, A., et al., Evaluating Diagnosis-Baded Case-MIx Measures: How well do they apply to the VA Population? Medical Care, 2001. **39**(7): p. 692-704.

Figures

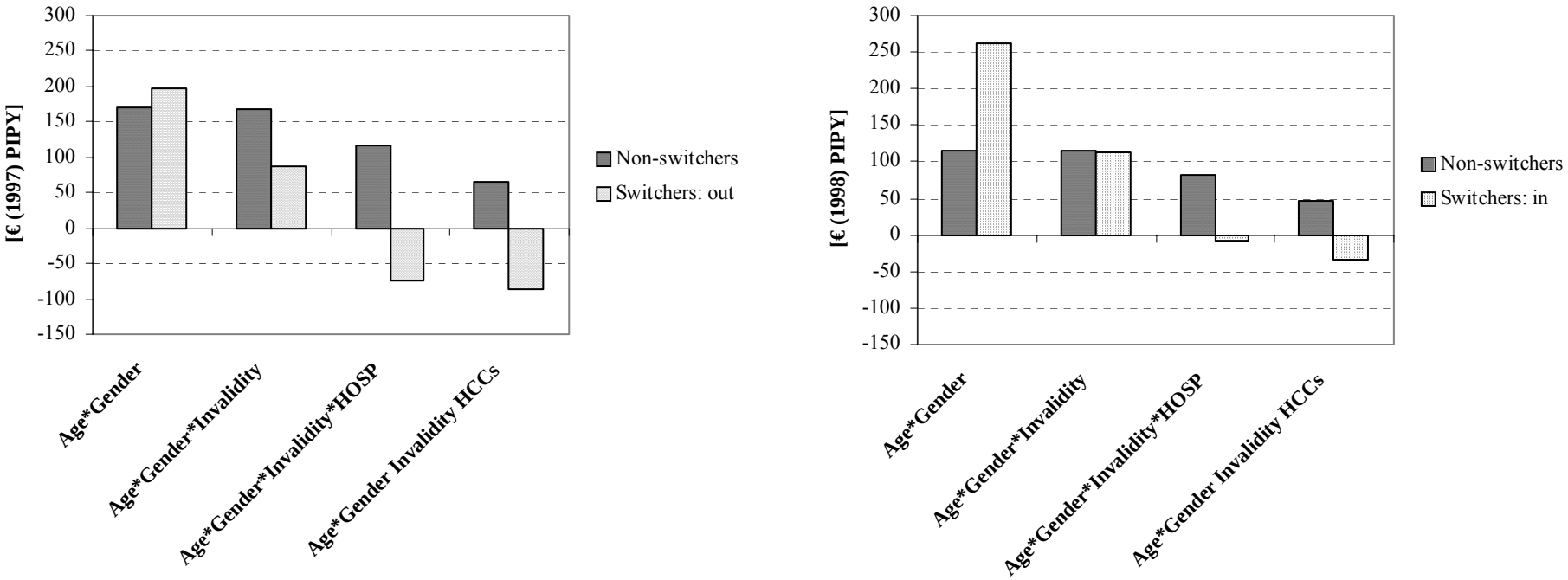

Figure 1: Predictable mean gains and losses for non-switchers, leaving switchers, and joining switchers by risk assessment model

Tables

Table 1: Demographics, expenditure and inpatient experience of the sickness fund population

	Concurrent application	Prospective application
N	788,130	733,378
Year(s)	1997	1997/1998
% female	51.4	51.9
Age		
Mean	45.4	45.7
% age under 18	17.2	17.1
% age 18-45	32.8	31.9
% age 46-65	25.5	26.3
% age over 65	24.5	24.7
Mean months of cover		
Year 1	11.4	11.7
Year 2	--	11.6
Year-1 expenditure		
Mean	€ 1,740	€ 1,643
Standard deviation	€ 5,106	€ 4,222
Coefficient of variation (x 100)	294	257
Median	€ 518	€ 523
Year-2 expenditure		
Mean	--	€ 1,762
Standard deviation	--	€ 5,256
Coefficient of variation (x 100)	--	298
Median	--	€ 513
% with non-zero expenditure		
Year 1	5.8	4.9
Year 2	--	4.7
% ever hospitalised		
Year 1	16.2	15.7
Year 2	--	17.2

Table 2: Summary of individual level predictive performance – 10-fold cross-validation mean

	Concurrent application			Prospective application		
N	78,813 (x 10)			73,337 (x 10)		
Mean expenditure	€ 1,740			€ 1,762		
Risk Adjustment Model	**MAPE [€]**	**$R^2_{adj.}$ [%]**	**CPM [%]**	**MAPE [€]**	**$R^2_{adj.}$ [%]**	**CPM [%]**
Age*Gender	1,842	3.2	6.7	1,902	3.1	6.3
Age*Gender*Invalidity	1,795	5.1	9.0	1,862	4.5	8.3
Age*Gender*Invalidity*HOSP	1,227	24.6	37.8	1,786	7.6	12.0
Offered weights DCG/HCC: Age*Gender Invalidity RRWs	1,171	31.2	40.7	1,763	10.3	13.2
Reparameterised DCG/HCC: Age*Gender Invalidity HCCs	1,067	37.3	45.9	1,738	11.7	14.4

Table 3: Summary of group level predictive performance for actual (non-annualised) expenditure quintiles – 10-fold cross-validation mean

		Concurrent application			Prospective application		
	Risk Adjustment Model	***Mean*** **MP [€]**	**MAPE [€]**	**PR**	***Mean*** **MP [€]**	**MAPE [€]**	**PR**
Q-1	***Actual***	***53***			***55***		
	Age*Gender	1,262	1,209	24.0	1,283	1,227	23.2
	Age*Gender*Invalidity	1,101	1,048	20.9	1,147	1,092	20.7
	Age*Gender*Invalidity*HOSP	490	439	9.3	1,008	953	18.2
	Age*Gender Invalidity RRWs	563	511	10.7	1,021	966	18.4
	Age*Gender Invalidity HCCs	463	412	8.8	981	926	17.7
Q-2	***Actual***	***220***			***229***		
	Age*Gender	1,273	1,054	5.8	1,362	1,134	5.9
	Age*Gender*Invalidity	1,158	939	5.3	1,265	1,037	5.5
	Age*Gender*Invalidity*HOSP	518	306	2.4	1,154	926	5.0
	Age*Gender Invalidity RRWs	586	373	2.7	1,144	917	5.0
	Age*Gender Invalidity HCCs	481	284	2.2	1,109	881	4.8
Q-3	***Actual***	***508***			***512***		
	Age*Gender	1,714	1,213	3.4	1,755	1,248	3.4
	Age*Gender*Invalidity	1,650	1,151	3.2	1,711	1,205	3.3
	Age*Gender*Invalidity*HOSP	779	399	1.5	1,594	1,095	3.1
	Age*Gender Invalidity RRWs	899	508	1.8	1,572	1,071	3.1
	Age*Gender Invalidity HCCs	749	417	1.5	1,536	1,041	3.0
Q-4	***Actual***	***1,185***			***1,151***		
	Age*Gender	2,128	1,120	1.8	2,092	1,115	1.8
	Age*Gender*Invalidity	2,203	1,262	1.9	2,160	1,230	1.9
	Age*Gender*Invalidity*HOSP	1,416	780	1.2	2,135	1,246	1.9
	Age*Gender Invalidity RRWs	1,483	775	1.3	2,081	1,195	1.8
	Age*Gender Invalidity HCCs	1,300	689	1.1	2,068	1,201	1.8
Q-5	***Actual***	***6,584***			***6,815***		
	Age*Gender	2,263	4,567	0.3	2,289	4,780	0.3
	Age*Gender*Invalidity	2,511	4,513	0.4	2,490	4,732	0.4
	Age*Gender*Invalidity*HOSP	5,384	4,147	0.8	2,878	4,690	0.4
	Age*Gender Invalidity RRWs	5,061	3,634	0.8	2,953	4,652	0.4
	Age*Gender Invalidity HCCs	5,594	3,482	0.8	3,077	4,625	0.5

Table 4: Summary of predictive performance with expenditure truncation at € 20,000 of actual (non-annualised) expenditure – individual results and results for actual (non-truncated) expenditure quintiles

Risk Adjustment Model	**MAPE [€]**	**$R^2_{adj.}$ [%]**	**CPM [%]**	**PR Q-1**	**Q-2**	**Q-3**	**Q-4**	**Q-5**
	Concurrent application							
N			788,130					
Mean expenditure			€ 1,613					
Age*Gender	1,646	6.0	7.4	22.0	5.3	3.1	1.7	0.4
Age*Gender*Invalidity	1,602	8.5	9.8	19.7	4.9	3.0	1.7	0.4
Age*Gender*Invalidity*HOSP	1,055	40.3	40.6	9.2	2.3	1.5	1.1	0.8
Reparameterised DCG/HCC: Age*Gender Invalidity HCCs	913	52.3	48.6	9.0	2.3	1.5	1.1	0.8
	Prospective application							
N			733,378					
Mean expenditure			€ 1,623					
Age*Gender	1,685	6.1	7.0	21.1	5.5	3.2	1.7	0.3
Age*Gender*Invalidity	1,648	8.0	9.0	19.2	5.1	3.1	1.7	0.4
Age*Gender*Invalidity*HOSP	1,578	12.5	12.9	17.1	4.7	2.9	1.7	0.4
Reparameterised DCG/HCC: Age*Gender Invalidity HCCs	1,539	16.0	15.0	16.7	4.6	2.8	1.7	0.4

Table 5: Summary of predictive performance with expenditure truncation at € 10.000 above predicted (non-annualised) expenditure – individual results and results for actual (non-truncated) expenditure quintiles

Risk Adjustment Model	**MPC [€]**	**MAPE [€]**	**$R^2_{adj.}$ [%]**	**CPM [%]**	**PR Q-1**	**Q-2**	**Q-3**	**Q-4**	**Q-5**
				Concurrent application					
N				788,130					
Age*Gender	1,499	1,473	8,1	8,1	20,4	5,0	2,9	1,6	0,4
Age*Gender*Invalidity	1,509	1,445	10.9	10.6	18.5	4.6	2.8	1.6	0.4
Age*Gender*Invalidity*HOSP	1,587	1,021	46.4	41.2	9.2	2.3	1.5	1.1	0.8
Reparameterised DCG/HCC: Age*Gender Invalidity HCCs	1,636	928	60.5	48.8	8.6	2.1	1.4	1.1	0.8
				Prospective application					
N				733,378					
Age*Gender	1,501	1,501	8.1	7.5	19.5	5.1	2.9	1.6	0.4
Age*Gender*Invalidity	1,509	1,476	10.3	9.7	17.9	4.8	2.9	1.6	0.4
Age*Gender*Invalidity*HOSP	1,524	1,432	15.7	13.6	16.2	4.5	2.7	1.6	0.4
Reparameterised DCG/HCC: Age*Gender Invalidity HCCs	1,542	1,413	22.8	16.1	15.7	4.3	2.7	1.6	0.5

Table 6: Retrospective actual and predicted mean expenditure, mean absolute prediction error, and ratio of predicted to actual expenditure for newborn and deceased leavers, stayers, and joiners by risk assessment model and reference model

Risk Adjustment Model	***Mean* MP [€]**	**MAPE [€]**	**PR**	***Mean* MP [€]**	**MAPE [€]**	**PR**
	Year: 1997			**Year: 1998**		
	Leavers deceased in 1997 (n = 13,863)					
Actual	***14,734***					
Age*Gender	2,969	12,514	0.20			
Age*Gender*Invalidity	3,178	12,382	0.22			
Age*Gender*Invalidity*HOSP	6,757	10,002	0.46			
Age*Gender Invalidity HCCs	9,918	8,845	0.67			
	Stayers deceased in 1998 (n = 13,753)					
Actual	***6,180***			***14,846***		
Age*Gender	2,963	5,206	0.48	3,018	12,664	0.20
Age*Gender*Invalidity	3,137	5,155	0.51	3,218	12,527	0.22
Age*Gender*Invalidity*HOSP	4,369	3,743	0.71	6,776	10,103	0.46
Age*Gender Invalidity HCCs	5,464	3,264	0.88	10,158	8,844	0.68
	Joiners born in 1998 (n = 3,784)					
Actual				***4,041***		
Age*Gender[#]				4,134	5,216	1.02
Age*Gender*HOSP[#]				4,123	3,636	1.02
Age*Gender HCCs[#]				4,135	3,084	1.02
	Stayers born in 1997 (n = 3,854)					
Actual	***3,658***			***1,692***		
Age*Gender[#]	3,816	4,669	1.04	1,077	1,704	0.64
Age*Gender*HOSP[#]	3,826	3,198	1.05	1,300	1,189	0.77
Age*Gender HCCs[#]	3,814	2,677	1.04	1,678	1,116	0.99

[#] *Entitlement to an invalidity pension is confined to individuals aged 18 to 65.*

Table 8: Summary of predictive performance for models incorporating an LTC indicator – individual results and results for actual (non-truncated) expenditure quintiles, overall samples

Risk Adjustment Model	**MAPE [€]**	**$R^2_{adj.}$ [%]**	**CPM [%]**	**PR Q-1**	**Q-2**	**Q-3**	**Q-4**	**Q-5**
			Concurrent application					
N			788,130					
Mean expenditure			€ 1,740					
Age*Gender*Invalidity	1,794	5.1	9.1	20.9	5.3	3.2	1.9	0.4
Age*Gender*LTC	1,788	5.6	9.4	22.5	5.5	3.2	1.8	0.4
Age*Gender Invalidity*LTC	1,764	6.3	10.6	20.6	5.1	3.1	1.8	0.4
Age*Gender*Invalidity*HOSP	1,226	24.6	37.9	9.3	2.4	1.5	1.2	0.8
Age*Gender*LTC*HOSP	1,228	24.7	37.8	9.7	2.4	1.5	1.2	0.8
Age*Gender Invalidity*LTC HOSP	1,329	22.9	32.6	6.8	1.6	1.5	1.4	0.8
Reparameterised DCG/HCC: Age*Gender Invalidity HCCs	1,066	37.6	46.0	8.8	2.2	1.5	1.1	0.8
Reparameterised DCG/HCC: Age*Gender LTC HCCs	1,069	37.6	45.8	9.5	2.3	1.5	1.1	0.8
Reparameterised DCG/HCC: Age*Gender Invalidity*LTC HCCs	1,058	37.9	46.4	8.6	2.1	1.4	1.1	0.9
			Prospective application					
N			733,378					
Mean expenditure			€ 1,762					
Age*Gender*Invalidity	1,862	4.6	8.3	20.7	5.5	3.3	1.9	0.4
Age*Gender*LTC	1,864	4.7	8.2	22.1	5.7	3.3	1.8	0.4
Age*Gender Invalidity*LTC	1,841	5.4	9.3	20.4	5.4	3.2	1.9	0.4
Age*Gender*Invalidity*HOSP	1,785	7.7	12.1	18.2	5.0	3.1	1.9	0.4
Age*Gender*LTC*HOSP	1,788	7.7	11.9	19.2	5.2	3.1	1.8	0.4
Age*Gender Invalidity*LTC HOSP	1,788	7.7	11.9	17.8	4.9	3.1	1.9	0.4
Reparameterised DCG/HCC: Age*Gender Invalidity HCCs	1,736	12.0	14.5	17.7	4.8	3.0	1.8	0.5
Reparameterised DCG/HCC: Age*Gender LTC HCCs	1,743	11.8	14.2	18.7	5.0	3.0	1.8	0.4
Reparameterised DCG/HCC: Age*Gender Invalidity*LTC HCCs	1,725	12.4	15.0	17.5	4.8	2.9	1.8	0.5

Diskussionsbeiträge (ab 1995)

85 **Kempa, B.; Nelles, M.; 1995:** Interest Rate Differentials in a Target Zone - Model with Heterogeneous Expectations. Essen

86 **Westerhoff, Dr. H.-D.; 1995:** Maximum-Likelihood Schätzer für Poisson Lags. Essen

87 **Brüggelambert, G.; 1995:** Determinanten der Erwartungs- und Entscheidungsbildung an politischen Börsen. Essen

88 **Schiess, H.-F.; 1995:** Die theoretische Relevanz der ECU für die externe Unternehmensfinanzierung. Essen

89 **Nibbrig, B.; 1995:** Anlagebetrug - Anlegerschutz - ein bedeutsames (noch weiterhin) vernachlässigtes Thema in der wirtschaftlichen Bildung. Essen

90 **von der Lippe, P.; 1995:** Statistische Wohlfahrtsindikatoren - Die Messung des Lebensstandards - Essen

91 **von der Lippe, P.; Zwick, M.; 1996:** Die Lebensqualität von Gemeinden. Einige Anmerkungen zum Lebensqualität-Atlas von Korczak aus der Sicht der Statistik. Essen

92 **von der Lippe, P.; 1996:** Szenarien für eine künftige EU-Intrahandelsstatistik. Essen

93 **Kladroba, A.; 1996:** Zeitvariable Parameterschätzungen mit „Flexible Least Squares“, Essen

94 **Spaetling, D.; 1996:** Anmerkungen zum Duopolgleichgewicht auf einem vollkommenen Markt bei isoelastischer Marktnachfrage. Essen

95 **Moritz, K-H. / Ohnsorge, F,; 1996:** Arbeitnehmermitbestimmung aus der Sicht der Property-Rigths-Theorie. Essen

96 **Moritz, K.-H.; 1996:** Komparativ-statische Analyse von Haushaltsgleichgewichten: eine alternative Vorgehensweise. Essen

97 **Brüggelambert, G.; 1996:** About Institutions and Efficiency in Political Stock Markets: Views from some Spin-Offs. Essen

98 **Lippe v.d., P.M.; 1997:** Die Rolle der amtlichen Statistik in einer Marktwirtschaft und die deutschen Erfahrungen beim Übergang zu einer marktwirtschaftlichen Statistik. Essen

99 **Lippe v.d., P.M.; 1997:** Einige Grundsätze der Preisstatistik und Preisindextheorie. Essen

100 **Lippe v.d., P.M.; 1997:** Preisindizes für Zwecke der Deflationierung (Arten der Deflationierung, indextheoretische Probleme). Essen

101 **Peschel, P.; 1997:** Arbeitsmarkterfahrungen junger Wirtschaftswissenschaftler - Ergebnisse der Essener Absolventenbefragung. Essen

102 **Jennes, B.; 1998:** Ein alternativer Indikator der deutschen Geldpolitik - Untersuchung im Rahmen eines VAR-Modells. Essen

103 **Lippe v.d., P.M.; 1998:** Die amtliche Statistik in Deutschland, kein zu privatisierender Betrieb. Essen

104 **Lippe v.d., P. M.; 1998:** The political role of official statistics in the former GDR (East Germany). Essen

105 **Lippe v.d., P. M.; 1998:** Zwei Beiträge zum Verhältnis zwischen „Statistik" und „Demokratie": „Einführung eines TES-Seminars" und „Die politische Rolle der amtlichen Statistik in der ehemaligen DDR". Essen

106 **Kladroba, A.; 1998:** Der Kalman-Filter bei fehlerhafter Hyperstruktur - eine Simulationsstudie. Essen

107 **Westerhoff, H.-D.; 1998**: Erweiterung und Reform der EU. Essen

108 **Dewenter, R.; 1999:** Messung von substitutionalen Beziehungen mithilfe von Kreuzkorrelationsfunktionen - Eine alternative Methode der Marktstrukturanalyse Essen

109 **Westerhoff, H.-D.; 1999:** Kondratieff-Zyklen - Zum Stand der Diskussion -

110 **von der Lippe, P.; 1999:** Volkswirtschaftliche Gesamtrechnung(en) in Deutschland nach neuen internationalen Vorschriften. Essen

111 **Kladroba, Andreas; 1999:** Schumacher oder Häkkinen: Wer war der „wahre" Formel 1 Weltmeister 1998? Einige Anmerkungen zum Aggregationsproblem bei der Erstellung von Rankings. Essen

112 **Jörges, Katharina; Süß, Stefan**; **2000:** Das idealtypische Menschenbild virtueller Unternehmen und seine Konsequenzen. Essen

113 **Nienhüser, Werner / Becker, Christina / Jans, Manuel; 2000:** Studentische Erwerbstätigkeit und Teilzeit-Studium – Erste Ergebnisse einer schriftlichen Befragung aller Studierenden der Wirtschaftswissenschaften an der Universität GH Essen

114 **Weinem, Michael / Carstens, Stefan; 2000:** Incentive Compatibility of Dual Transfer Pricing

115 **Brüggelambert, Gregor; 2000:** Nachteile von Gründungen in jungen Märkten beim Umgang mit Behörden und bei der Kreditaufnahme. Essen

116 **Nienhüser, Werner; 2001:** Studentische Erwerbstätigkeit und Probleme im Studium - Eine empirische Analyse geschlechtsspezifischer Unterschiede. Essen

117 **Kladroba, Andreas / von der Lippe, Peter; 2001:** Die Qualität von Aktienempfehlungen in Publikumszeitschriften. Essen

118 **Westerhoff, Horst-Dieter; 2002:** Globalisierung und Bilanzrecht – Einige ordnungspolitische Bemerkungen. Essen

119 **Clausen, Volker / Wohltmann, Hans-Werner; 2002:** Monetary and Fiscal Policy Dynamics in an Asymmetric Monetary Union. Essen

120 **Hayo, Bernd / Seifert, Wolfgang; 2002:** Subjective Economic Well-being in Eastern Europe. Essen

121 **Hayo, Bernd / Doh Chull Shin; 2002:** Popular Reaction to the Intervention by the IMF in the Korean Economic Crisis. Essen

122 **Jans, Manuel; 2002:** Überlegungen zur Güte der transaktionskostenökonomischen Erklärung betrieblicher Personalstrategien. Essen

123 **Kladroba, Andreas; 2002:** Die Qualität von Aktienempfehlungen in Publikumszeitschriften (Teil 2). Essen

124 **Wohltmann, Hans-Werner / Clausen, Volker; 2002:** Foreign Price Shocks and Monetary Policy in an Asymmetric Monetary Union. Essen

125 **Westermann, Michael / Dewenter, Ralf; 2003:** Cinema Demand in Germany. Essen

126 **Von der Lippe, Peter; 2003:** Zur Interpretation gesamtwirtschaftlicher Produktions- und Einkommensgrößen der deutschen Volkswirtschaftlichen Gesamtrechnung nach dem SNA 93/ESVG 95. Essen

127 **Von der Lippe, Peter; 2003:** Das Identifikationsproblem in der Ökonometrie. Essen

128 **Jans, Manuel; 2003:** Sozialkapitalkonzepte und ihre Brauchbarkeit in der Personal- und Organisationsforschung. Essen

129 **Kempa, Bernd; 2003:** An oversimplified inquiry into the sources of exchange rate variability. Essen

130 **Wasem, Jürgen / Greß, Stefan / Hessel, Franz; 2003**: Hausarztmodelle in der GKV - Effekte und Perspektiven vor dem Hintergrund nationaler und internationaler Erfahrungen. Essen

131 **Kladroba, Andreas; 2003**: Das neue Datenschutzrecht in der betrieblichen Praxis. Essen

132 **Emami Namini, Julian; 2004:** International business cycle transmission in a dynamic multi-sectoral Heckscher-Ohlin model

133 **Westermann, Michael / Krol, Bianca; 2004:** Zur Wirksamkeit von Practice Tests am Beispiel der Klausur Deskriptive Statistik

134 **Behrend, Corinne / Buchner, Florian / Happich, Michael / Holle, Rolf / Reitmeir, Peter / Wasem, Jürgen; 2004:** Risk-Adjusted Capitation Payments: How Well Do Principal Inpatient Diagnosis-Based Models Work in the German Situation? Results From a Large Data Set