

Improving risk adjustment without diagnostic information – Evidence from reforms in Switzerland

Reto Bürgin

bure@zhaw.ch

ZHAW Zurich University of Applied Sciences

Lukas Kauer

University of Lucerne

Michael Stucky

ZHAW Zurich University of Applied Sciences

Christina Vetsch-Tzogiou

ZHAW Zurich University of Applied Sciences

Research Article

Keywords: risk equalization, pharmacy-based cost groups (PCG), health spending, capitation payments

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11039815/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

A major goal of risk adjustment (RA) schemes in health insurance systems is to mitigate insurers' incentive for risk selection. While most countries incorporate diagnostic information into these schemes, Switzerland has faced unique challenges due to limited diagnostic information. To account for morbidity in the outpatient sector, the Swiss RA scheme first introduced a cost-based indicator (spending threshold on drugs) in 2017, which was replaced by a utilization-based indicator using pharmaceutical cost groups (PCG) in 2020.

Objectives

The aim of this study was to compare the performance of these two different risk adjusters, to assess their potential impact on the incentives for risk selection, and to discuss the potential policy implications.

Methods

For the first time in the evaluation of the Swiss RA scheme, the entire adult population (6.99 m in 2022) was included. We examined individual-level spending predictions, quantifying over- or undercompensation in four distinct RA models, and explored whether the current RA scheme continues to present possibilities for risk selection by investigating spending variation among similarly compensated individuals.

Results

Both indicators substantially increased the goodness of fit and led to a considerable reduction in under- and overcompensation in many subgroups. However, several of the currently used PCGs exhibited significant spending variation which may offer new potential for risk selection.

Conclusion

Even with limited diagnostic information, an RA scheme can be substantially improved by including morbidity information based on drug prescriptions.

JEL I13, I11, I18, G22

Background

Risk adjustment (RA) in health plan payment is widely used to reduce the financial risk caused by the different risk profiles of the insured and to mitigate risk selection incentives for the insurer (McGuire & Van Kleeef, 2018; van de Ven et al., 2023). Effective RA design identifies reliable predictors of future high spending which is then compensated through risk equalization. Most countries (e.g., the Netherlands, Germany, USA) use diagnostic codes (e.g., based on ICD-10) in their RA schemes. In some countries, like Switzerland, diagnostic information is not available to the insurer. Cost-based or utilization-based indicators are popular alternatives, each with distinct benefits and limitations. In fact, there is an inherent trade-off between fit and power (Douven et al., 2026; Geruso & McGuire, 2016). A high fit of an indicator, which means that it explains a lot of variation in spending, is usually related to low power, a reduced incentive for the insurer to provide services efficiently. On top of that, gaming incentives can be introduced if the insurer can manipulate the indicator to trigger a higher RA payment. This problem is similar to the upcoding problem in the payment of health care providers.

Countries like the Netherlands run sophisticated RA schemes relying on several risk indicators based on diagnosis, spending, utilization, and socioeconomic background of enrollees (van Kleeef et al., 2018). This requires substantial data transfers between the insurer and the regulator. Switzerland has a political system based on federalism and local-first governance (Article 5a of the Swiss constitution). Highly sensitive diagnostic information is not transferred from health care provider to payer, which can be the insurer or the state. The use of diagnostic information in the RA scheme has therefore never been an option. Instead, from its introduction in 1994 until 2012, the Swiss RA scheme only relied on demographic information (age and sex).

The first risk indicator addressing morbidity was an indicator for an inpatient stay in the previous year. A person's spending is predicted to be significantly higher if they had an inpatient stay (mainly at a hospital or a nursing home) of three consecutive nights in the previous year (Holly et al. (2004), Beck et al. (2006)). While this indicator is relatively simple to collect, it has two major limitations: its dichotomy and its focus on morbidity related to inpatient treatment, with conditions treated in an outpatient setting not being captured.

In a major reform in 2014, the RA scheme became a permanent institution and the decision to which indicators should be included in the RA shifted from the parliament to the Ministry of Health (Federal Office of Public Health, FOPH). Aware of the deficiencies of the current RA scheme, the FOPH addressed the missing outpatient sector by including a threshold of drug spending of CHF 5000 ($\approx$ € 4500 in 2017) in the previous year starting in 2017. The threshold of CHF 5000 was chosen as a simulation showing this number would maximize fit (Pirktl, 2015). Enrolees with drug spending above this threshold were grouped into a separate risk class and split depending on whether the inpatient stay indicator also applied to them. Again, this indicator was relatively simple to implement but created new incentives for gaming and reduced the incentive for efficiency. Since compensation for a person above the threshold was very high, the incentive for the insurer to increase the spending of a person to reach the threshold was high. One example is that the incentive to provide generics instead of the brand drug was reduced, as higher drug prices increased the probability to crossing the threshold. This is why the FOPH planned early on to eventually replace the threshold with pharmaceutical cost groups (PCG) in 2020.

This concept was developed in the US and implemented in the Netherlands (Lamers & Vliet, 2003). Despite its name, it is a utilization-based indicator where enrollees need to cross a certain threshold of so called defined daily doses (DDD) of a group of drugs to qualify for a PCG. In this way, the price and brand of

the drug do not matter. Because certain drugs are only used for conditions that are persisting and linked to expected high future costs, they can be combined into groups and serve as a proxy for a diagnosis. The regulator can regularly adapt the number of PCGs, the qualifying drugs and DDD thresholds. In 2022, 34 PCGs were included in the Swiss RA scheme.

The risk equalization system in Switzerland has the same formula but is run separately in all 26 cantons. It is a zero-sum transfer among insurers (i.e. it follows an internal system or Modality B in the terminology of van de Ven and Ellis (2000). Insurers collect a monthly community-rated premium from their enrolees. An insurer then receives (pays) a contribution for every enrolee with costs above (below) the cantonal average. These contributions are summed up for every insurer so that some insurers are net payers, while some are net recipients. In 2022, 20 of the 45 insurers were net payers.

Our study is the first to use data on the full population and over multiple years (2020–2022) to assess the performance of different RA schemes in Switzerland. The many evaluations of the Swiss RA scheme since 1998 were usually limited to a single insurer or small number of insurers or a geographically limited region (Schmid et al., 2018). The most recent PCG reform was prospectively analysed by Kauer et al. (2020). With data from one insurer and a different list of PCGs, they showed that the new model substantially improved the compensation of very expensive enrolees, but at the same time led to many highly under- and overcompensated individuals.

The goal of this paper is to evaluate performance and impact on incentives for insurers of two risk adjusters which measure morbidity not through diagnostic codes but through drug spending and utilization, respectively. We show that the cost-based indicator is easy to implement, substantially increases fit but is susceptible to gaming and power incentives. The utilization-based indicator increases fit even more and is less susceptible to gaming. However, it is more complex and requires more effort to implement burdening the insurers. The Swiss experience to introduce a crude indicator while awaiting the development of more refined indicators can be helpful for other countries that are in the process of setting up or improving their RA schemes.

Methods

Data

Two unique anonymized individual-level datasets were used: 1) on the RA calculation from the “Gemeinsame Einrichtung KVG” (2020–2022), the institution which runs the annual risk equalization in Switzerland, and 2) on the mandatory health insurance (2020–2021) by the FOPH. Both data sets cover the insured adult (19+) population (6.99 m in 2022) as children are excluded from RA. The datasets are collected routinely and in a uniform way by all insurers such that, within the dataset, individuals can be linked over time even if they change their insurer or their insurance plan. These datasets have so far not been used for any empirical evaluation of the RA. Our results are based on the most recent year of implementation for which data is available, which corresponds to the RA 2022 if only dataset 1 is required, and to RA 2021 otherwise.

Dataset 1 includes all variables needed for the calculation of the RA, i.e., total net health care spending, age, sex, indicator for an inpatient stay for at least three consecutive nights in the previous year, as well as information on all drugs consumed in the previous year (at the level of specific packages), from which we constructed the PCGs along with the drug spending threshold. Dataset 2 includes information on the insurance plan (e.g., managed care plan), the yearly deductible (CHF 300, 500, 1000, 1500, 2000, 2500), the region of residence, the premium, whether the individual received premium subsidies by the government, as well as the duration of insurance (< 12 months in case of death or migration).

Comparison of RA model performance 2014–2020

We compared four RA models that show the RA reform steps taken between 2014 and 2020 Table 1 Table 1. To enhance comparability, the models were modified compared to how they were used in practice. Specifically, for the two outdated models (M1, M2) we implemented the current prospective calculation scheme that takes into account claims from a longer period (26 instead of 14 months) and that integrates a price inflation to adjust spending from the year RA is calculated to the year RA is applied to. For the current RA model with PCGs, we disregarded the reform introduced in 2019 to redistribute funds between the young and the adult population. The purpose of this reform was to reduce the transfers from the young to the old and any effect on the fit of the RA model was unintended. Including model M4 allowed to analyse to what extent the risk adjusters in M2 and M3 are supplements as predictors. For a detailed description of the Swiss RA model, we refer to the according regulation 832.112.1 available in French, German or Italian or (2020) available in German.

Since there is no consensus in the literature which goodness of fit measures to report (van Veen et al., 2015), models are compared along a comprehensive set, including measures that have been used to evaluate RA models in the past (adjusted R^2 , Cummings Prediction Measure (CPM), mean absolute error (MAE), median absolute deviation (MAD)), and common measures for statistical model selection (Akaike information criterium (AIC), Bayes' information criterium (BIC)).

The estimation error is defined as the difference between estimated net spending minus observed net spending in the year of adjustment, in CHF per person per month. As the RA is calculated for each of the 26 Swiss cantons separately, the estimation error is equal to the cantonal average of net spending plus the RA payment minus the observed net spending. A positive estimation error indicates overcompensation, a negative estimation error indicates undercompensation. Because the estimated net costs are calculated based on the net spending of the year before the adjustment, the estimation error can be interpreted as an “out-of-sample error”. Consequently, the computed goodness of fit indicators do not (necessarily) improve with additional risk adjusters and the sum of the estimation errors does not (necessarily) sum up to zero.

Table 1
Risk adjustment (RA) models evaluated

RA Model	Description
M1	Base model (RA model from 2014 plus 26 months billing period and price inflation)
M2	M1 + high drug spending (RA model from 2017)
M3	M1 + PCG (RA model from 2020 without lower transfers from young adults)
M4	M3 + high drug spending (model introduced for analytical reasons only)
<i>Notes: Prospective calculation means that the costs in the previous year are adjusted to the equalization year using a price inflator.</i>	

Assessment of incentives for risk selection

The six statistical goodness of fit measures show how precisely the RA model estimates the net spending per month in the equalization year. As the major goal of RA is keeping risk selection incentives low, it is also important to assess the quality of the model regarding selection incentives. We therefore also examined the extent to which insured individuals were under- or overcompensated according to their spending in the previous year. The weaker the under- or overcompensation effects along the spending in the previous year, the better the model quality regarding selection incentives.

Besides this analysis at the aggregate level, we analysed the differences in the mean estimation errors between groups of insured in the various models. These groups are defined based on variables that are observed by the insurer (from the previous year) but not included in the model, thus offering potential for risk selection (net spending of > CHF 1000 /month, multiple PCGs, deductible level, accident coverage, insurance plan type, means-tested premium subsidy). Accident coverage reflects non-employment, as the employed population usually has accident insurance coverage through their employer. Receiving a premium subsidy by the state is a proxy for low income.

In a next step, we evaluated the distribution of estimation errors within each of the 34 PCGs in the RA model currently in use. PCGs with a large variation of estimation errors could potentially be used in the selection of low-cost individuals within risk-adjusted groups. PCGs in which estimation errors are distributed unimodally and symmetrically around zero are likely to offer less potential for risk selection. We therefore compared the mean, median, and standard deviation of the estimation error between PCGs.

Assessment of impacts on insurance premiums

Higher fit should compress the premium distribution in the market. We evaluated the impact on the insurer market by assessing the mean premiums by type of insurer and the variation in offered premiums before and after the reforms (2017–2021). A reduction in the variation coupled with a higher increase in mean premiums in insurers with a healthier risk pool (net payers) indicate a reduction in risk selection incentives. To this end, we used the premiums for each enrollee (from dataset 2; enrollees with accident insurance coverage only). We computed the mean premium for each insurer and compared the change in mean premiums over time by type of insurer. Insurers that were net payers or net recipients in every year, with average per-insured-month amounts into/out of the RA scheme exceeding CHF 20 (absolute) in at least three years, were assigned to the respective category (payers or recipients). Insurers with average per-insured-month (absolute) amounts below CHF 20 in every year were classified as neutral. Insurers falling into different categories across years were assigned to the most frequent one. Insurers assigned to every category at least once over the five years were labelled neutral.

In addition, we assessed the variation in mean premiums over time. To make the variation in premiums comparable across years, we calculated the coefficient of variation. This measure is unitless and is calculated as the standard deviation divided by the mean. We assessed the variation at the level of premium regions and computed a weighted average of the coefficient of variation using the sum of insurance months by premium region and year. Results

Model performance

Table 2 shows the various performance measures for the four different RA models for the adjustment year 2022. The RA model's ability to predict net spending improved substantially with the new indicators. All six model performance measures improved from M1 to M2 and further to M3. For instance, the MAE decreased from CHF 382 to CHF 326. The improvement in model performance from M1 to M2, reflecting the inclusion of the indicator for high drug spending in the previous year, was about half as large as the improvement from M1 to M3, reflecting the inclusion of PCGs. The net volume redistributed among insurers increased by 24% from M1 (CHF 1.4 bn) to M3 (CHF 1.7 bn), while the volume redistributed among all insured persons (including those within the same insurer) increased by even 33%, from CHF 10.3 bn in M1 to CHF 13.7 bn in M3. To put these numbers into context, the total net spending borne by health insurers in 2022 amounted to CHF 32.8 bn. The performance of models M3 and M4 is very similar, suggesting that combining PCGs with the indicator for high drug spending does not provide a substantial complementary benefit. Additional analyses in which the considered performance measures were calculated for the years 2020 and 2021 indicated that the results are robust over time.

Table 2
Comparison of RA model performance (RA-year 2022).

Goodness of Fit indicators	M1	M2	M3	M4
Adjusted R^2	0.15	0.24	0.31	0.33
CPM	0.17	0.23	0.30	0.30
MAD (CHF)	199	175	145	147
MAE (CHF)	382	346	326	322
AIC/10 ⁶	94.49	93.10	92.96	92.78
BIC/10 ⁶	94.51	93.12	92.98	92.81
Redistributed volume				
Between insurers (bn CHF)	1.4	1.8	1.7	1.7
Between insured persons (bn CHF)	10.3	11.8	13.7	13.6
N (m)	6.99	6.99	6.99	6.99
<i>Notes: Dependent variable is net spending in subsequent year. For the definition of M1-M4, see Table 1.</i>				
<i>CPM: Cummings Prediction Measure, MAD: median absolute deviation, MAE: mean absolute error, AIC: Akaike information criterium, BIC: Bayes' information criterium</i>				

Incentives for risk selection within specific subgroups of the population

Risk selection may be performed at the group level and not the entire population. Therefore, we analysed the mean estimation errors for subpopulations defined by a set of variables that are not included in the RA model.

Figure 1 compares the mean estimation errors by groups of net monthly health care spending in the previous year for the four RA models. Positive values refer to an overcompensation, negative values refer to an undercompensation. Again, the more sophisticated risk indicator (M3, M4) performs best in that the overcompensation for the low spenders as well as the undercompensation for the high spenders are at the lowest in almost every cost group. The most striking difference between the four models is in the highest spending group ($> =$ CHF 1,000/month).

Figure 1: Overcompensation/undercompensation in models M1-M4 by monthly net spending in previous year (RA-year 2022).

Notes: Values above 0 refer to an overcompensation. The grey bars represent the number of enrollees in the respective group (right hand scale).

The deductible is a proxy for self-assessed health, as healthier enrollees tend to choose higher deductibles. In both M1 and M3, the lower deductible was associated with an undercompensation, and the higher deductible was associated with an overcompensation. However, individuals with a lower deductible were much less undercompensated in M3 than in M1, and individuals with a higher deductible were much less overcompensated in M3 than in M1, respectively. This shows that M3 substantially decreased the difference in the average over-/undercompensation between the deductible levels (see Figure A1 in the supplementary information).

The results for other grouping variables such as having multiple PCGs, accident coverage, the type of insurance plan, and receiving premium subsidy (a proxy for low income) are shown in Table 3. The compensation of net spending was improved for all grouping variables with the implementation of M3. The most substantial improvements were found for net spending $>$ CHF 1000/month in the previous year, multiple PCGs in the previous year, and accident coverage in the previous year. The improvement with respect to those having multiple PCGs is the least surprising as the PCGs are explicitly incorporated into M3. Similar patterns were observed for the sum of absolute estimation errors, which declined markedly across all grouping variables. Despite these improvements, notable estimation errors remained for some groups, particularly multiple PCGs, indicating that even the enhanced model does not fully eliminate predictable under- or overcompensation in these populations.

Table 3
Maximum group differences in mean estimation errors and sum of absolute estimation errors (RA-year 2021)

Group variable	Mean absolute estimation errors (CHF/month and enrollee)				Sum of absolute estimation errors in billion CHF per year			
	M1	M2	M3	M4	M1	M2	M3	M4
Net spending > CHF 1000/month in previous year	1031.7	569.9	501.8	451.7	6.1	4.3	3.0	2.6
Reduction MX – M1 (%)		44.8%	51.4%	56.2%		29.9%	50.8%	56.7%
Deductible in previous year	209.0		129.5		4.7		2.3	
Reduction M3 – M1 (%)			38.0%				49.8%	
Multiple PCG in previous year	765.9	450.3	41.3	28.1	3.3	2.9	0.2	0.1
Reduction MX – M1 (%)		41.2%	94.6%	96.3%		13.1%	94.6%	96.4%
Accident coverage in previous year	95.5		48.6		1.7		0.1	
Reduction M3 – M1 (%)			49.1%				94.9%	
Insurance plan in previous year	96.6		58.5		2.0		0.5	
Reduction M3 – M1 (%)			39.4%				75.1%	
Premium subsidy in previous year	130.5		80.6		2.1		0.5	
Reduction M3 – M1 (%)			38.2%				77.1%	
<i>Notes: Some values could not be calculated for M2 due to missing data.</i>								

New risk selection incentives within PCGs

Several PCGs showed a high systematic variation in the estimation error (see Table 4). This may create new incentives for risk selection on previously unattractive enrollees compared with the RA models used before the introduction of PCGs. Insurers may benefit from attracting individuals with one of these PCGs whose actual spending is below the predictable level. The PCG “kidney diseases” (NIE) was the most heterogeneous PCG with respect to the estimation error. It showed the highest mean overcompensation (CHF 358.5/month). The median was negative (CHF – 251.7/month), indicating that most individuals within that PCG were undercompensated. Other heterogeneous PCGs include the PCG “cystic fibrosis” (ZFP), “pulmonary arterial hypertension” (PAH) and “complex cancer” (KRK).

Table 4
Distribution of estimation errors (CHF/month and insured) within PCG based on RA model M3 (using data from 2022)

PCG	Estimation error (RA model M3)				
	PCG name	Insured persons	Mean	Median	Std. Err.
NIE	Kidney diseases	4,641	358.5	-251.7	4,893.7
KRE	Cancer	647	201.9	715.4	1,689.6
MSK	Multiple Sclerosis	12,522	100.0	252.4	1,240.6
HIV	HIV / AIDS	14,227	55.7	261.9	1,494.6
TRA	Transplantations	7,389	50.3	560.9	3,047.8
PSY	Psychosis	46,560	42.6	368.1	1,349.8
AIK	Autoimmune disorders	38,770	39.1	253.5	1,354.5
KRK	Complex cancer	36,800	36.6	824.6	3,950.9
ADH	ADHD	19,634	20.7	219.8	1,058.9
DEP	Depression	363,111	7.9	289.3	1,285.3
MCR	Morbus Crohn / ulcerative colitis	11,228	7.9	280.0	1,250.3
BSR	Bipolar disorders (regular)	8,113	7.3	281.8	1,282.0
KHO	Hormone-sensitive tumors	30,615	7.0	318.0	1,272.4
SMC	Chronic pain without opioids	93,982	6.6	261.3	1,315.8
DM2	Diabetes mellitus type 2	67,875	5.6	220.5	1,175.9
ABH	Addiction without Nicotine	5,279	4.1	332.2	1,434.7
SMN	Neuropathic Pain	21,264	4.0	488.0	2,061.9
DM2 + hyp	Diabetes mellitus type 2 & hypertension	134,656	0.5	252.7	1,315.4
DM1	Diabetes mellitus type 1	64,322	-4.3	373.1	1,759.9
THY	Thyroid diseases	146,926	-5.1	185.3	1,176.2
RHE	Rheumatism	19,119	-9.8	247.3	1,289.1
HCH	High cholesterol	509,670	-10.1	188.7	1,196.1
COP	COPD / severe asthma	59,034	-10.8	476.6	1,927.4
EPI	Epilepsy	50,173	-12.4	364.6	1,555.9
AST	Asthma	74,732	-14.9	231.5	1,363.7
GLA	Glaucoma	120,672	-16.1	202.6	1,203.5
CAR	Heart diseases	44,563	-16.1	354.9	1,769.3
PAR	Parkinson's	16,141	-32.6	424.1	1,722.4
ZNS	Diseases of the central nervous system	8,581	-35.8	706.2	2,526.0
PSO	Psoriasis	7,465	-37.4	260.2	1,270.6
ALZ	Alzheimer's	12,092	-69.0	189.8	1,391.0
WAS	Growth disorder	153	-74.3	272.4	1,858.8
PAH	Pulmonary arterial hypertension	960	-75.5	913.8	5,111.8
ZFP	Cystic fibrosis	4,536	-226.8	1,494.6	5,387.1

The distributions of estimation errors for these four PCGs are shown in Fig. 2. The distribution was left-skewed in all PCGs, due to a few insured people with very high net spending per month and thus high negative estimation errors. The PCG "kidney diseases" (NIE) showed a bimodal distribution. The PCG "complex cancer" (KRK) had a tendency towards a bimodal distribution, but to a lesser extent.

Impact on insurance premiums

The analysis of mean insurance premiums before and after the reform (2017–2021) indicate a convergence of premiums across insurers. First, insurers with a more favourable risk pool before the reform in 2020 (net payers) increased their insurance premiums more (+ 9.8%) than insurer with a less favourable risk

pool (net recipients; +4.8%) between 2017 and 2021. In 2020, the year of the RA reform, mean premiums increased only in net payers (+1.7%), but decreased in insurers classified as neutral (-0.6%) and in recipients (-0.1%). Second, the variation of premiums offered on the market converged: The coefficient of variation decreased in 35 out of 42 premium regions (83%) between 2019 and 2020. The weighted average of the coefficient of variation was 13.8% in 2017, 12.5% in 2019 and 11.1% in 2021. Detailed results on the mean premiums by insurer type and the premium variation by region across time are shown in the appendix (Tables A2 and A3).

Discussion

Comparing the performance of different RA models using a variety of statistical measures allowed us to evaluate the effect of single reform steps between 2014 and 2020 in the Swiss RA. Our results show that including only a simple measure of high drug use, such as drug spending above CHF 5000 per year, improves model performance substantially. The performance improves even further using a more sophisticated set of 34 PCGs. Using both indicators together does not substantially improve the performance suggesting no complementary benefits from the drug indicator when combined with PCGs.

Comparing our fit measures to previous literature in Switzerland summarized in Schmid et al. (2018, Table 16.2), results are roughly similar although the R^2 in this study is slightly lower for the drug threshold (0.24 vs. 0.25/0.28), while it is slightly higher for the PCGs (0.30 vs. 0.21/0.24/0.25/0.30). All these previous analyses are based on data from a limited number of insurers, while we had access to the full population. Furthermore, previous analyses of the PCGs worked with a different set of PCGs than the ones which were finally implemented.

Comparing R^2 internationally can be treacherous as the spending an insurer is responsible for can vary substantially between countries. For example, Switzerland has significant cost sharing from enrolees (copayment) and cantons (in inpatient care). Nevertheless, comparing our results to van Kleef et al. (2013) (2013) who perform a similar analysis in the Netherlands, their reported jump in R^2 from a demographic-based RA model to one including PCGs from 0.06 to 0.16 is similar in size.

Since risk selection might not be performed at the individual level but on the group level, we calculated over- and undercompensation for various groups. The two drug-based risk adjusters improved both measures with PCGs again outperforming the drug threshold. However, the distribution of the estimation error may not always be homogenous among all PCGs. Compared to a unimodal distribution, a bimodal or multimodal distribution within a PCG that is included in RA increases the incentives for risk selection. In our data, only the PCG “kidney diseases” (NIE) showed a clear bimodal distribution of the estimation errors. This is also visible to some extent in “complex cancer” (KRK). For the remaining PCGs, the distributions were found to be unimodal. However, it remains unclear whether it is possible to identify any grouping variable that allows for the distinction of the two modal values and to operationalize it for risk selection. One discriminating factor could be disease severity. As the treatment of end-stage kidney disease involves dialysis, health care spending of the severely ill may be substantially higher than of patients in earlier disease stages. It may therefore be useful to collect data on dialysis from all insured individuals and include this as an additional variable in the RA model.

The potential for risk selection is high whenever there are substantial differences between over- and undercompensated individuals. This may not only be the case with bi- or multimodal distributions of estimation errors, but also with unimodal distributions with large variation/a high coefficient of variation. In addition to the distribution form of the estimation errors, the level of the PCG contribution (benefit for individuals with this PCG) and the prevalence of PCG are also important. The higher the PCG contribution and the higher the prevalence, the more attractive the PCG for risk selection and gaming.

Kauer et al. (2020) simulate the impact of the introduction of PCGs in Switzerland on the tails of the residual spending distribution prior to their introduction. With a limited sample and different PCG definitions, they show similar patterns in that undercompensation is significantly reduced but also creating a new group of highly overpaid within PCGs.

A central insight from the literature on health-plan payment design is that risk adjusters must be evaluated not only on statistical fit but also on their incentive properties, particularly power (the share of marginal costs borne by the insurer) and gaming/manipulability (the scope for strategic behaviour that increases payments without improving efficiency). Geruso and McGuire (2016) operationalize power via counterfactual simulation on claims data: impose a small exogenous reduction in utilization or costs for a sample of enrolees, recompute the full payment and calculate how much of the cost reduction the insurer keeps. Because our analysis relies on aggregate evaluation data rather than individual claims, we cannot quantify the exact change in power or the magnitude of potential gaming. Future work with micro-level claims would allow simulation of marginal cost retention and direct measurement of threshold-crossing elasticities. Still, we can provide a qualitative assessment. Because PCGs are typically triggered by sustained or clinically meaningful drug regimens rather than marginal increments, they tend to preserve higher average power than more granular utilization-based mechanisms while still improving compensation for high-morbidity groups. In contrast, a pure drug-thresholds adjuster introduces a discrete notch in the payment schedule. Below the threshold, power remains close to that of a purely prospective system; once the threshold is crossed, payment rises sharply, sharply reducing marginal reimbursement for additional drug costs. This creates uneven power across the distribution of drug spending: high power for low- and moderate-cost patients and low power (or even negative marginal incentives) near and above the threshold. The resulting imbalance can distort clinical decisions, encouraging insurers to manage care differently depending on where patients sit relative to the cutoff. Consequently, while both mechanisms reduce power relative to a demographic-only model, threshold-based adjusters are likely to produce greater dispersion in service- or patient-specific power and therefore weaker overall cost-control incentives.

Regarding gaming incentives, their nature and severity differ between the two adjusters. PCGs reward the pattern of drug use consistent with specific chronic conditions. Gaming would require sustained changes in prescribing, which are constrained by clinical guidelines, patient need, and potential scrutiny from regulators or physicians. Compared to these more graduated incentives, the blunt all-or-nothing payoff structure of the drug threshold amplifies gaming potential much more.

Moreover, it is important to consider the costs of implementing the reform itself. The RA reforms came with substantial costs for the institutions involved in the RA, mainly the insurers. A recent survey among health insurers showed that the time and effort needed in the implementation of the RA reforms and the data requirements have caused a substantial additional burden for insurers (Bürgin et al., 2024). This underlines that no reform comes without costs, and that these costs for the roughly 40 private insurers should be taken into account in future reforms. Several reactions on the insurance market indicate a decrease in the incentives for risk selection triggered by the RA reforms. Our analysis of the mean health insurance premiums and its variation before and after the reform showed that there was a convergence in the offered premiums between 2017 and 2021. This is likely to be caused by the improved compensation for sicker populations. Premium variation across insurers decreased and insurers which were recipients in the RA before the reform (i.e., had a more morbid insured population) showed below-average premium growth rates. On the other hand, net payers had to increase their premiums disproportionately as the RA contributions increased substantially after the reform. Furthermore, the popular risk selection strategy of creating conglomerates by establishing subsidiaries to circumvent the community-rated premium regulation (von Wyl & Beck, 2016) is no longer attractive. By 2024, eleven subsidiaries were merged with the parent company and the number of conglomerates decreased from eight to two (Bürgin et al., 2024).

Policy considerations and implications

Reforming Swiss RA has been a time-consuming process. Compared to other countries with risk equalization among insurers, the Swiss RA was quite simple in the beginning and left substantial room for risk selection (von Wyl & Beck, 2016). As this paper shows, the inclusion of information on drugs used in outpatient settings improved RA considerably.

RA in other countries such as the Netherlands or Germany use diagnostic information. Diagnostic codes are currently not available for RA in Switzerland. While the recent Swiss reforms demonstrate the substantial improvements in RA enabled by incorporating PCGs, further improvements of a similar magnitude seem unlikely in the near future without the use of explicit diagnostic information.

Limitations and future research

This analysis has several limitations. First, the evaluation period coincided with the COVID-19 pandemic, which may have influenced health care spending and the performance of the current RA scheme. Second, we were only able to observe incentives for risk selection but not the actual behaviour of the insurers. Whether or not insurers engage in risk selection could not be investigated in this study. However, our result on the premium convergence is a strong indication that risk selection behaviour has decreased. Third, while our data are comprehensive including the full population, the information on enrolees remains limited. Over- and undercompensation could only be calculated for a limited set of groups. The lack of claims data also disallowed the mentioned power analysis above.

Future research should therefore repeat the analysis in few years using data from years not affected by the pandemic and after insurers have adapted to the new RA scheme. Furthermore, it should evaluate if insurers are trying to use the new potential for risk selection enabled by large spending variations within PCGs or to game the system by upcoding (somehow pushing their enrolees over the PCG (DDD) threshold to receive the compensation for the PCG).

Conclusion

Two recent reforms of the Swiss RA scheme have led to substantial improvements of the equalization scheme. In lack of diagnostic information in the outpatient sector, Switzerland focused on outpatient drug spending. Both, a simple, dichotomous risk indicator based on a threshold of high drug spending and even more so, a more sophisticated indicator based on specific drug utilization have helped reduce the incentives for risk selection. Policy makers face a trade-off: A cost-based, easy to implement indicator, yet with lower power and a higher incentive for gaming versus a utilization-based, more complicated to implement indicator with higher power and a lower incentive for gaming. Switzerland chose the latter, most likely because the additional fit from it is still substantial. However, we show that this new indicator may lead to new potentials for risk selection on previously unattractive enrolees. We therefore recommend a close monitoring of the insurers' behaviour.

Declarations

Ethical Approval

Ethical approval was not required for this study as it involved the secondary analysis of fully anonymized individual-level data.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Author Contribution

R.B. contributed to conceptualization, methodology, project administration, data curation, formal analysis, investigation, software development, visualization, and writing (review and editing). M.S. contributed to conceptualization, methodology, project administration, formal analysis, writing (original draft), and writing (review and editing). C.V.-T. contributed to conceptualization, methodology, writing (original draft), and writing (review and editing). L.K. contributed to

conceptualization, methodology, supervision, visualization, and writing (review and editing). All authors reviewed the manuscript and approved the final version.

Acknowledgement

The authors would like to thank Andreas Kohler, Anna Drewek, Christoph Thommen, Marcel Dettling, and Simon Wieser for their valuable contributions to the project that formed the basis of this research. The authors also gratefully acknowledge the contributions and expert advice provided by Lennart Pirkli and Richard C. van Kleef. Their insights and feedback were highly valuable to the development of the project and the present study. We are grateful to Thomas G. McGuire and participants of the Swiss Health Economics Workshop 2024, of the Risk Adjustment Network meeting 2025, and of the International Health Economics Association congress 2025 for helpful comments.

Data Availability

The anonymized individual-level datasets used for the risk adjustment (RA) calculations were obtained from the Gemeinsame Einrichtung KVG, which administers the Swiss mandatory health insurance risk equalization scheme. In addition, individual-level data on mandatory health insurance were obtained from the Federal Office of Public Health (FOPH). These data were used under licence for the current study and are not publicly available. Access to these datasets is subject to approval by the respective data holders and applicable legal and data protection requirements.

References

1. Beck, K., Trottmann, M., Käser, U., Keller, B., von Rotz, S., Zweifel, P.: Nachhaltige Gestaltung des Risikoausgleichs in der Schweizer Krankenversicherung. Ott (2006)
2. Bürgin, R.: *Berechnungsformeln für den Risikoausgleich mit PCG ab 2020*. Bundesamt für Gesundheit BAG. (2020). https://www.bag.admin.ch/dam/bag/de/dokumente/kuv-aufsicht/pus/risikoausgleich/calculcdr-pcginclus-2020.pdf.download.pdf/200420_KUV%20VA_Berechnungsformeln%20f%C3%BCr%20den%20Risikoausgleich%20mit%20PCG%20ab%202020_V6_DE.pdf
3. Bürgin, R., Stucki, M., Vetsch-Tzogiou, C., Kauer, L., Kohler, A., Drewek, A., Thommen, C., Dettling, M., Wieser, S.: *Wirkungsanalyse zum Risikoausgleich mit pharmazeutischen Kostengruppen (PCG). Schlussbericht im Auftrag des Bundesamts für Gesundheit (BAG)*. (2024)
4. Douven, R., Brammli, Shuvi, Buchner, F., Kauer, L., van Kleef, R., McGuire, T.G., Paolucci, F., Reitsma, M., Rose, S., Schmid, C.P.R., Wasem, J.: Power of volume and price incentives in health plan payment models: A country comparison. *EsCHER Working Paper No. 2026003*. (2026). <https://www.eur.nl/en/media/2026-04-escher-2026003>
5. Geruso, M., McGuire, T.G.: Tradeoffs in the design of health plan payment systems: Fit, power and balance. *J. Health. Econ.* **47**, 1–19 (2016). <https://doi.org/10.1016/j.jhealeco.2016.01.007>
6. Holly, A., Gardiol, L., Eggli, Y., Yalcin, T., Ribiero, T.: Health-Based Risk Adjustment: An Exploration Using Medical Information from Prior Hospitalisation. Institut d'économie et management de la santé (IEMS), University of Lausanne (2004)
7. Kauer, L., McGuire, T.G., Beck, K.: Extreme under and overcompensation in morbidity-based health plan payments: The case of Switzerland. *Health Policy*. **124**(1), 61–68 (2020). <https://doi.org/10.1016/j.healthpol.2019.11.008>
8. Lamers, L.M., Vliet, R.C.J.A.: Health-based risk adjustment Improving the pharmacy-based cost group model to reduce gaming possibilities. *Eur. J. Health Economics: HEPAC: Health Econ. Prev. Care*. **4**(2), 107–114 (2003). <https://doi.org/10.1007/s10198-002-0159-9>
9. McGuire, T.G., Van Kleef, T.G.: Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets. Academic (2018)
10. Pirkli, L.: *Verfeinerung des Risikoausgleichs durch Berücksichtigung der Arzneimittelpreise*. (2015). <https://www.bsv.admin.ch/dam/bsv/de/dokumente/themenuebergreifend/publikationen/CHSS/chss-01-2015.pdf.download.pdf/de-01-2015.pdf>
11. Schmid, C.P.R., Beck, K., Kauer, L.: Chapter 16—Health Plan Payment in Switzerland. In T. G. McGuire & R. C. van Kleef (Hrsg.), *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets* (S. 453–489). Academic Press. (2018). <https://doi.org/10.1016/B978-0-12-811325-7.00016-6>
12. van de Ven, W., Ellis, R.P.: Risk Adjustment in Competitive Health Plan Markets*. In A. J. Culyer & J. P. Newhouse (Hrsg.), *Handbook of Health Economics* (Bd. 1, S. 755–845). Elsevier. (2000). [https://doi.org/10.1016/S1574-0064\(00\)80173-0](https://doi.org/10.1016/S1574-0064(00)80173-0)
13. van de Ven, W., Hamstra, G., van Kleef, R., Reuser, M., Stam, P.: The goal of risk equalization in regulated competitive health insurance markets. *Eur. J. Health Econ.* **24**(1), 111–123 (2023). <https://doi.org/10.1007/s10198-022-01457-7>
14. van Kleef, R.C., Eijkenaar, F., van Vliet, R.C.J.A., van de Ven, W.P.M.M.: Chapter 14—Health Plan Payment in the Netherlands. In T. G. McGuire & R. C. van Kleef (Hrsg.), *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets* (S. 397–429). Academic Press. (2018). <https://doi.org/10.1016/B978-0-12-811325-7.00014-2>
15. Van Kleef, R.C., Van Vliet, R.C., Van de Ven, W.P.: Risk equalization in The Netherlands: An empirical evaluation. *Expert Rev. Pharmacoeconomics Outcomes Research*. **13**(6), 829–839 (2013). <https://doi.org/10.1586/14737167.2013.842127>
16. van Veen, S.H.C.M., van Kleef, R.C., van de Ven, W.P.M.M., van Vliet, R.C.J.A.: Is there one measure-of-fit that fits all? A taxonomy and review of measures-of-fit for risk-equalization models. *Med. Care Res. Review: MCRR*. **72**(2), 220–243 (2015). <https://doi.org/10.1177/1077558715572900>
17. von Wyl, V., Beck, K.: Do insurers respond to risk adjustment? A long-term, nationwide analysis from Switzerland. *Eur. J. Health Econ.* **17**(2), 171–183 (2016). <https://doi.org/10.1007/s10198-015-0669-x>

Figures

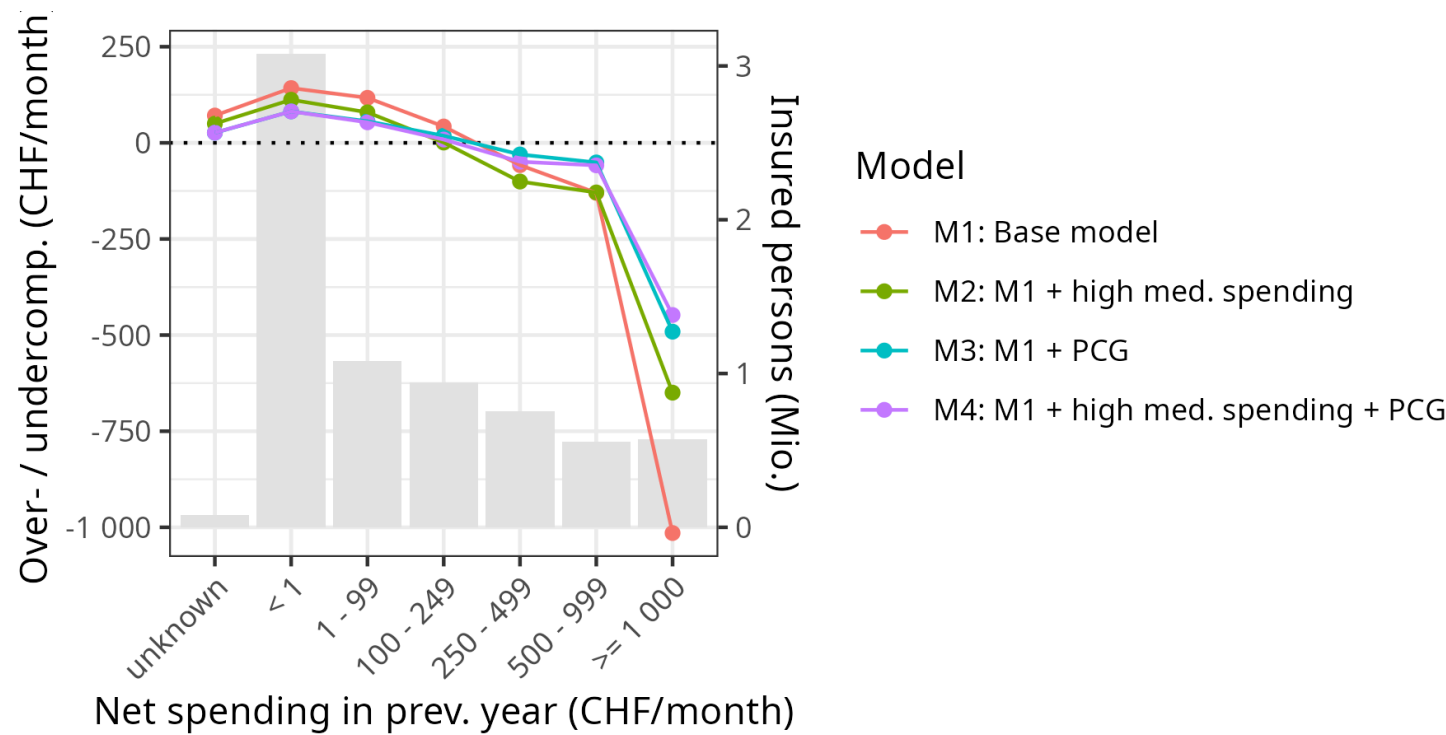


Figure 1

Overcompensation/undercompensation in models M1-M4 by monthly net spending in previous year (RA-year 2022).

Notes: Values above 0 refer to an overcompensation. The grey bars represent the number of enrolees in the respective group (right hand scale).

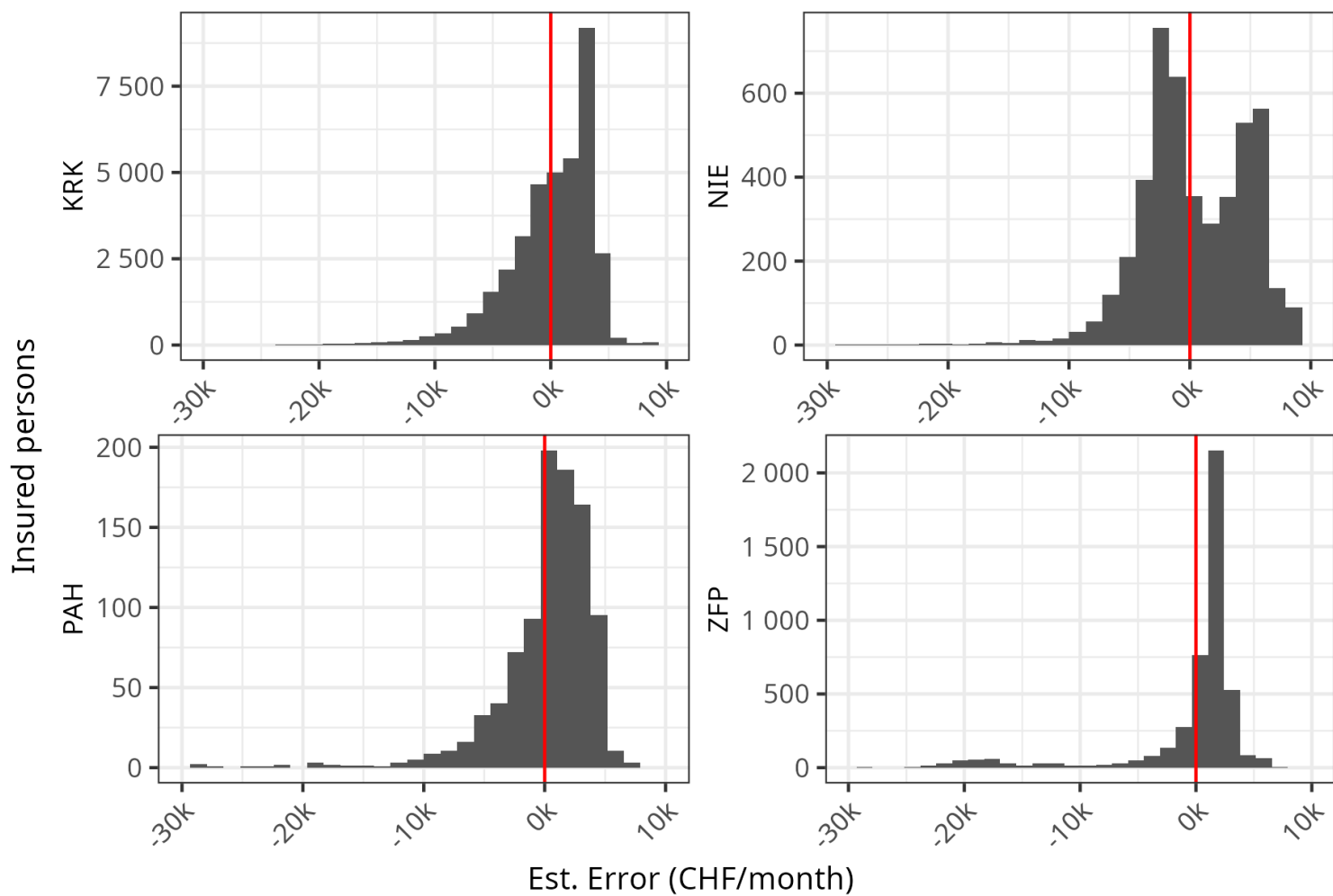


Figure 2

Distribution of estimation error for four selected PCGs with heterogeneous error distribution (RA model M3, RA-year 2022).

Notes: KRK: complex cancer, NIE: kidney diseases, PAH: pulmonary arterial hypertension, ZFP: cystic fibrosis

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [RiskAdjPCGEuJoHeEcSupInfSub1.docx](#)